Instituições de acolhimento

U.PORTO

FMUP FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO

UNIDADE LOCAL DE SAÚDE
SÃO JOÃO

Artigo 48º, § 3º

A Faculdade não responde pelas doutrinas expendidas na Dissertação.

(Regulamento da Faculdade de Medicina do Porto – Decreto-Lei nº 19337, de 29 de janeiro de 1931)

Dissertação de candidatura ao grau de Doutor apresentada à Faculdade de Medicina da Universidade do Porto, no âmbito do Programa Doutoral em Neurociências Clínicas, Neuropsiquiatria e Saúde Mental

Título *(Title)*: Os resultados reportados pelo doente em patologia degenerativa da coluna vertebral – rumo a ferramentas válidas e fiáveis para avaliar o sucesso cirúrgico
Patient-reported outcomes in degenerative spine surgery – towards valid and reliable tools to assess surgical success

Candidato *(PhD student)*: Pedro Daniel Valente Aguiar

Orientador *(supervisor)*: Prof. Doutor Paulo Miguel da Silva Pereira
Professor Associado Convidado da Faculdade de Medicina da Universidade do Porto

Coorientador *(co-supervisor)*: Prof. Doutor Paulo José Campos Linhares Vieira
Professor Auxiliar Convidado da Faculdade de Medicina da Universidade do Porto

TABLE OF CONTENTS

I. ACKNOWLEDGMENTS

Like any research project, this PhD thesis required a great deal of study, dedication and research that translated into hundreds of hours spent away from those who matter most, my family and friends.

To my wife, Paulina, and my daughter, Liliana - the emotional backbone of my life - that give me the strength and motivation to keep improving everyday.

To my parents for all the love, confidence and support over the years.

To the rest of my wonderful family and friends.

To my patients.

To my mentors and colleagues.

II. LIST OF PUBLICATIONS (Art. 8º do Decreto-lei nº 388/70)

This thesis was based on the following articles:

1 - Valente Aguiar, P., Silva, P. S., Vaz, R., & Pereira, P. (2023). Are the results of patient-reported outcome measures after spine surgery influenced by recall of preoperative scores? A randomized controlled trial. *The Spine Journal, 23*(3), 369–378. https://doi.org/10.1016/j.spinee.2022.11.007

2 - Valente Aguiar, P., Santos Silva, P., Lucas, D., Vaz, R., Pereira, P., & Mannion, A. F. (2024). Cross-cultural adaptation, validation and establishment of the minimal clinically important change score of the European Portuguese Core Outcome Measures Index in patients with lumbar degenerative spine disease. *European Spine Journal, 33*(2), 394–400. https://doi.org/10.1007/s00586-023-08093-6

3 - Valente Aguiar, P., Pereira, P., Mannion, A. F., & Santos Silva, P. (2025). Cross-cultural adaptation, validation, and establishment of the minimal clinically important change score of the European Portuguese Core Outcome Measures Index in patients with cervical degenerative spine disease. *European Spine Journal, 34*(1), 36–42. https://doi.org/10.1007/s00586-024-08564-4

III. LIST OF ABBREVIATIONS

COMI - Core Outcome Measures Index

ESAS - Edmonton Symptom Assessment System

ePROMs – electronic Patient-Reported Outcome Measures

EQ-5D - EuroQol Five Dimension Questionnaire

ES – Effect size

HAQ-DI - Health Assessment Questionnaire Disability Index

ICC - Intraclass correlation coefficients

MCIC - Minimal clinically important change

MCID - Minimal clinically important difference

MDC – Minimal detectable change

NDI - Neck Disability Index

NRS - Numeric Rating Scale

ODI - Oswestry Disability Index

PCQH - Pharos Checklist for Questionnaires in Healthcare

PROMs - Patient-Reported Outcome Measures

RCT – Randomized controlled trial

ROC - Receiver operating characteristic

RMDQ - Roland-Morris Disability Questionnaire

SF-36 - Short Form 36 Health Survey Questionnaire

SD – Standard deviation

SEM – Standard error of measurement

VAS - Visual Analogue Scale

IV. SUMMARY

Background

Degenerative spine disorders are a leading cause of disability worldwide, often requiring surgical intervention when conservative treatments fail. As populations age and life expectancy increases, the prevalence of these conditions continues to rise, placing a growing burden on individuals, healthcare systems, and economies. However, defining surgical success remains challenging due to the subjective nature of symptoms associated with degenerative spine disease. Traditional outcome metrics—such as radiographic findings or surgeon-reported measures—often fail to capture the full scope of a patient's postoperative experience. This has led to a paradigm shift toward incorporating the patient's voice into outcome evaluation, with patient-reported outcome measures (PROMs) now widely recognized as indispensable tools in both clinical practice and research. Given the central role that PROMs play in evaluating surgical outcome through a patient perspective, it is essential that these tools are both clinically meaningful and psychometrically valid within the context in which they are used.

Despite their widespread adoption, several limitations persist, including recall bias, varying cultural validity of tools, and the need for clearly defined clinically meaningful change thresholds. Therefore, there is a growing need for validated, reliable, and culturally adapted PROMs in spine surgery, especially in the European Portuguese-speaking population.

Aims

The main aim of this thesis was to contribute to the improvement of outcome assessment in degenerative spine surgery by enhancing the validity, reliability, and interpretability of patient-reported outcome measures.

Firstly, the thesis sought to investigate the extent and clinical relevance of recall bias in PROM data collection. In many clinical settings, baseline PROMs are not recorded prospectively, and patients are instead asked to retrospectively estimate their preoperative symptoms. This retrospective assessment may be influenced by the patient's memory bias, introducing a systematic error that could distort the perceived effectiveness of surgery. The thesis aimed to ascertain the impact of this bias on postoperative PROM scores and assess its implications for outcome interpretation.

Secondly, the research aimed to create and validate the European Portuguese versions of the Core Outcome Measures Index (COMI) for both patients with degenerative disease of the

lumbar and cervical spine. The COMI is a simple and efficient questionnaire that has been validated to multiple languages and is the PROM of choice for the European Spine Society. As such, the validation of the COMI in European Portuguese represents an important milestone for spine care in Portugal.

Finally, this thesis aimed to establish the minimal clinically important change (MCIC) for both the COMI-back and COMI-neck. By defining these thresholds, the thesis aimed to provide clinicians and researchers with robust criteria to evaluate the effectiveness of surgical interventions in a way that is both standardized and patient centered.

Together, these three aims address some critical gaps in the current use of PROMs in spine surgery and provide a foundation for more accurate, meaningful, and culturally relevant outcome measurement.

Methods

To address the objectives of the thesis, three complementary studies were conducted.

The first study was a randomized controlled trial designed to evaluate the impact of recall bias in PROMs. Patients scheduled for lumbar and cervical spine surgery were randomized to a control and an intervention group. The control group completed postoperative PROMs without having access to their preoperative scores in the same questionnaires. Patients in the intervention group were given their preoperative answers during fulfilment of postoperative PROMs.

The primary endpoint of the study was the median postoperative COMI score. The median postoperative Neck Disability Index (NDI) / Oswestry Disability Index (ODI) and EuroQol Five Dimension Questionnaire (EQ-5D) scores were the secondary endpoints.

For the COMI, NDI and ODI lower scores (range 0 to 10 for the COMI; 0 to 100 for NDI / ODI) mean better clinical status while for the EQ-5D lower scores (maximum score:1) mean worse clinical status.

The second and third studies focused on the cross-cultural adaptation and psychometric validation of the European Portuguese versions of the COMI-back and COMI-neck. Following international guidelines for translation and cultural adaptation, both instruments underwent forward and backward translation, expert review, and pilot testing. Subsequently, a cohort of patients undergoing surgery for lumbar or cervical degenerative spine conditions completed the adapted PROMs at different time points. Psychometric evaluation included assessments of reproducibility and test-retest reliability, construct validity, responsiveness, and floor/ceiling effects.

Finally, the thesis employed an anchor-based approach to establish the MCIC values for both European Portuguese versions of the COMI-back and COMI-neck.

Results

The findings of this thesis are based on the results of the three studies.

In the randomized controlled trial valid postoperative PROM data was available for 147 patients (median age 58 years; 59.9% female). At a median follow-up of 17 months, patients in the intervention group reported significantly better postoperative outcomes compared to the control group. This included lower scores in the COMI (median 4.20 vs. 5.45, $p = 0.023$), ODI/NDI (28.45 vs. 36.00, $p = 0.006$), and higher scores in the EQ-5D (0.689 vs. 0.586, $p = 0.017$). Because these findings were due to cervical patients a subgroup analysis of this cohort ($n = 66$) was performed. Only the cervical intervention group achieved a COMI improvement exceeding the minimal clinically important difference (MCID) (3.4 vs. 1.2) and had significantly better NDI and EQ-5D scores.

In the study validating the COMI-back, 108 (43 males, 65 females; mean age 56.4 years) patients completed the European Portuguese version of the COMI questionnaire preoperatively between January and December 2022. The mean baseline COMI score was 7.6, with sciatica (69.4%) and neurogenic claudication (28.7%) being the most common presenting symptoms. Psychometric analysis revealed no significant floor or ceiling effects. Construct validity was supported by excellent correlation between the COMI pain domain and Visual Analogue Scale (VAS) pain scores ($\rho = 0.85$), along with good correlations for function ($\rho = 0.49$), quality of life ($\rho = 0.54$) and disability ($\rho = 0.58$) when compared to the ODI and EQ-5D. The total COMI score also demonstrated strong correlation with the total ODI ($\rho = 0.63$) and EQ-5D ($\rho = -0.70$) scores. Test–retest reliability, assessed in 93 clinically stable patients, showed excellent intraclass correlation coefficients for the domains of well-being (0.84), quality of life (0.81), disability (0.86), and for the total COMI (0.89) score. The minimal detectable change was 1.52, and the MCIC was determined to be 2.1, with a sensitivity of 69% and specificity of 94%.

In the COMI-neck validation study, the questionnaires of 101 patients undergoing surgery for degenerative cervical spine disease (mean age 57.9 years; 51 males, 50 females) were analyzed. Myelopathy (73.3%) and radiculopathy (26.7%) were the most common diagnoses. The mean preoperative COMI score was 7.1. No significant floor or ceiling effects were observed, and construct validity was confirmed through strong correlations between COMI domains and comparator instruments: pain with VAS ($\rho = 0.95$), function and disability with the NDI ($\rho = 0.54$ and 0.47), and quality of life and well-being with EQ-5D ($\rho = -0.44$ and

−0.40). The total COMI score correlated very well with the NDI (ρ = 0.67) and well with the EQ-5D (ρ = −0.55). Reproducibility and test–retest reliability were assessed in 72 clinically stable patients over a 5-to-15-day interval, revealing stable scores (7.1 vs. 7.2; $p = 0.88$) and excellent ICCs for total COMI score (0.87), pain (0.90), and disability (0.85), with good reliability for function (0.78), well-being (0.79), and quality of life (0.74). The minimal detectable change was calculated at 1.76 points, while the minimal clinically important change was established at 2.0, with a sensitivity of 66% and specificity of 100%.

Conclusion

This thesis demonstrates that recall bias can affect patient-reported outcome scores following degenerative spine surgery, with patients given access to their baseline responses reporting significantly better outcomes—particularly among those undergoing cervical procedures.

In addition, the European Portuguese versions of the COMI-back and COMI-neck were successfully validated, showing strong construct validity, excellent reliability, and no significant floor or ceiling effects. Minimal clinically important change thresholds were also established, providing a benchmark for meaningful improvement in clinical practice. Together, these studies address key limitations in the use of PROMs in spine surgery, contributing validated, interpretable, and culturally adapted tools to support more reliable and patient-centered outcome assessment.

V. RESUMO

Introdução

A patologia degenerativa da coluna é uma das principais causas de incapacidade a nível mundial, frequentemente exigindo intervenção cirúrgica quando ocorre falência do tratamento conservador.

Com o envelhecimento populacional e o aumento da esperança média de vida a prevalência destas doenças tem aumentado, causando impacto crescente nos doentes, nos sistemas de saúde e na economia.

No entanto, a definição de sucesso cirúrgico nestes doentes é desafiante devido à natureza subjetiva da sintomatologia da patologia degenerativa raquidiana. As ferramentas tradicionalmente utilizadas para avaliar os resultados cirúrgicos tais como estudos imagiológicos ou medidas reportadas pelos cirurgiões frequentemente não permitem uma avaliação completa do estado clínico pós-operatório do doente. Isto levou a uma mudança de paradigma direcionado a incorporar a visão dos doentes para avaliar o resultado cirúrgico através das medidas reportadas pelos doentes (PROMs). Uma vez que as PROM são atualmente fundamentais na prática clínica e em investigação para avaliar o resultado cirúrgico através da perspetiva do doente, é essencial que estas ferramentas sejam clinicamente relevantes e possuam características psicométricas válidas dentro do contexto da sua utilização.

Apesar da ampla aplicação das PROM, subsistem várias limitações associadas às mesmas: viés de memória, diferenças na validade cultural e a necessidade de definir valores-limiar para mudanças clinicamente significativas. Assim, surge a necessidade de criar PROMs válidas, fiáveis e adaptadas culturalmente em doentes portugueses com patologia degenerativa da coluna vertebral.

Objetivos

O principal objetivo desta tese foi contribuir para a melhoria da avaliação dos resultados cirúrgicos em patologia degenerativa da coluna vertebral, assegurando a validação, fiabilidade e interpretabilidade das PROM.

Em primeiro lugar esta tese pretendeu investigar a extensão e relevância clínica do viés de memória nos resultados das PROM. Em vários contextos clínicos as PROM são recolhidas retrospetivamente e não de forma prospetiva, com os doentes a realizarem uma estimativa do seu estado clínico pré-operatório baseada na memória desse estado clínico. Esta análise

retrospetiva pode ser influenciada pelo viés de memória introduzindo um erro sistemático suscetível de influenciar a perceção da eficácia da cirurgia. Esta tese pretendeu avaliar o impacto deste viés nos resultados das PROM pós-operatórias e perceber o seu efeito na interpretação destes resultados.

Em segundo lugar pretendeu-se criar e validar para português europeu a versão do Core Outcome Measures Index (COMI) para doentes com patologia degenerativa da coluna lombar e cervical. O COMI é um questionário simples e eficaz que foi validado para múltiplas línguas e é a PROM de referência da European Spine Society. Desta forma, a validação do COMI para português europeu representou um marco importante para a avaliação da patologia de coluna vertebral no contexto nacional.

Finalmente, esta tese pretendeu definir a mudança mínima clinicamente importante (MCIC) para o COMI-lombar e COMI-cervical. Através da definição deste valor visou-se providenciar aos clínicos e investigadores critérios robustos, padronizados e focados no doente para avaliar a eficácia das intervenções cirúrgicas.

Em conjunto estes três objetivos visam atenuar algumas das limitações atuais na aplicação das PROM na cirurgia raquidiana e ser um alicerce para uma aplicação mais precisa, válida e culturalmente relevante destas ferramentas.

Métodos

Para se cumprirem os objetivos da tese realizaram-se três estudos complementares.

O primeiro foi um estudo randomizado controlado que pretendeu avaliar o impacto do viés de memória no resultado das PROM. Doentes a aguardar cirurgia por patologia degenerativa lombar e cervical foram randomizados para um grupo de controlo e outro de intervenção. O grupo de controlo completou as PROM pós-operatórias sem acesso aos resultados pré-operatórios enquanto os doentes no grupo de intervenção preencheram as PROM pós-operatórias com acesso às respostas pré-operatórias.

O desfecho primário do estudo foi o valor mediano do questionário COMI. Os desfechos secundários foram o valor mediano dos seguintes questionários: Neck Disability Index (NDI) / Oswestry Disability Index (ODI) e EuroQol Five Dimension Questionnaire (EQ-5D).

Para o COMI, NDI e ODI valores mais baixos (intervalo 0 a 10 para o COMI e 0 a 100 para NDI e ODI) representam melhor estado clínico enquanto para o EQ-5D valores mais baixos (resultado máximo:1) traduzem pior estado clínico.

O segundo e terceiro estudos focaram-se na adaptação transcultural e na validação psicométrica da versão para português europeu do COMI-lombar e COMI-cervical. A tradução

e adaptação cultural seguiram diretrizes internacionais com a realização de tradução direta, tradução reversa, avaliação por peritos e estudo piloto. A avaliação psicométrica incluiu avaliações da reprodutibilidade e fiabilidade teste-reteste, responsividade e efeitos de piso e teto.

Finalmente, estabeleceu-se a mudança mínima clinicamente importante do COMI-lombar e COMI-cervical através de um método de ancoragem.

Resultados

Os resultados desta tese baseiam-se nos resultados dos três estudos.

No estudo randomizado controlado foram avaliados 147 questionários válidos (idade mediana: 58 anos; 59,9% do sexo feminino). Os doentes no grupo de intervenção obtiveram resultados pós-operatórios significativamente melhores nas PROM face ao grupo de controlo após uma mediana de seguimento de 17 meses. Isto traduziu-se em valores mais baixos no COMI (mediana 4,20 vs. 5,45, p = 0,023), ODI/NDI (28,45 vs. 36,00, p = 0,006), e EQ-5D (0,689 vs. 0,586, p = 0,017). Uma vez que estes resultados se deveram aos doentes com patologia cervical fez-se uma análise deste subgrupo (n=66). Apenas os doentes do grupo de intervenção atingiram uma melhora clínica superior à diferença mínima clinicamente importante (MCID) no COMI-cervical (3,4 vs. 1,2), e os resultados do NDI e EQ-5D também foram significativamente melhores.

No estudo de validação do COMI-lombar, 108 doentes (43 homens, 65 mulheres; idade média 56,4 anos) completaram no pré-operatório a versão em português europeu deste questionário entre janeiro e dezembro de 2022. O valor médio do COMI foi 7,6 e a ciática (69,4%) e a claudicação neurogénica (28,7%) foram os principais sintomas de apresentação.

A análise psicométrica não revelou significativos efeitos de teto e piso e a validade de construto foi confirmada pelas seguintes correlações entre domínios: excelente entre a dor do COMI e da escala analógica da dor (VAS) ($\rho = 0,85$), boa entre função ($\rho = 0,49$), qualidade de vida ($\rho = 0,54$) e incapacidade ($\rho = 0,58$) quando comparados com o ODI e EQ-5D. O valor total do COMI demonstrou forte correlação com o valor total do ODI ($\rho = 0,63$) e EQ-5D ($\rho = -0,70$).

A reprodutibilidade e a fiabilidade teste-reteste foram avaliadas em 93 doentes clinicamente estáveis e verificaram-se excelentes coeficientes de correlação para os seguintes domínios do COMI: bem-estar (0,84), qualidade de vida (0,84), incapacidade (0,81) e valor total do questionário (0,89).

A variação mínima detetável foi de 1,52, e a mudança mínima clinicamente importante foi determinada como 2,1, com uma sensibilidade de 69% e especificidade de 94%.

No estudo de validação do COMI-cervical 101 doentes (51 homens, 50 mulheres; idade média 57,9 anos) completaram no pré-operatório a versão em português europeu deste questionário. O valor médio do COMI foi 7,1 e a mielopatia (73,3%) e a radiculopatia (26,7%) foram os sintomas mais frequentes.

Não se verificaram efeitos de teto ou piso significativos e a validade de construto foi confirmada com fortes correlações entre os domínios do COMI com outros questionários: dor com o VAS ($\rho = 0,95$), função e incapacidade com o NDI ($\rho = 0,54$ e $0,47$), qualidade de vida e bem-estar com o EQ-5D ($\rho = -0,44$ e $-0,40$).

O valor total do COMI correlacionou-se muito bem com o NDI ($\rho = 0,67$) e bem com o EQ-5D ($\rho = - 0,55$). A reprodutibilidade e a fiabilidade teste-reteste avaliaram-se em 72 doentes clinicamente estáveis num intervalo entre 5 e 15 dias e revelaram resultados semelhantes entre avaliações: (valor total 7,1 vs. 7,2; $p = 0,88$). Os coeficientes de correlação foram excelentes para o valor total do COMI (0,87), dor (0,90) e incapacidade (0,85) e foram bons para a função (0,78), bem-estar (0,79) e qualidade de vida (0,74).

A variação mínima detetável foi de 1,76, e a mudança mínima clinicamente importante foi determinada como 2, com uma sensibilidade de 66% e especificidade de 100%.

Conclusão

Esta tese demonstra que o viés de memória pode afetar os resultados das PROM após cirurgias para tratamento de patologia degenerativa da coluna. Os doentes com acesso às respostas pré-operatórias reportaram resultados significativamente melhores, particularmente os doentes com patologia cervical.

Adicionalmente, as versões para português europeu do COMI-lombar e COMI-cervical foram validadas com sucesso, e demonstraram uma boa validade de construto, excelente fiabilidade e ausência de relevantes efeitos de piso e teto. Foram ainda definidos limiares para a mudança mínima clinicamente importante, proporcionando um ponto de referência para se estabelecer benefício clínico com as intervenções cirúrgicas nestes questionários.

Em conjunto, estes estudos abordaram importantes limitações na utilização das PROM na cirurgia de coluna, contribuindo com ferramentas válidas, interpretáveis e culturalmente adaptadas para permitirem uma avaliação dos resultados cirúrgicos mais fiável e centrada no doente.

1. INTRODUCTION

1.1 Patient-reported outcome measures (PROMs) in clinical practice

1.1.1 Historical context and definition of PROMs

It was in the early 20th century that Dr. Ernest Amory Codman first envisaged the necessity for an outcome and patient-driven assessment of surgical results. [1]

This vision contributed to the transformation in health and spine care that has taken place since the 1970s with the realization that patient-reported outcome measures are fundamental tools in clinical practice and research in providing better patient care. [2]

Patient-reported outcome measures are questionnaires used to evaluate patient well-being, clinical status and treatment outcomes. Commonly evaluated domains include patient satisfaction with surgical or conservative outcome and medical care, symptoms such as pain / discomfort, functional status and overall health-related quality of life. [2,3]

Additionally, these tools provide clinicians the opportunity to compare outcomes between different health-care providers or procedures, help prioritize surgical candidates and can help identify areas of patient care that may have been overlooked during the treatment process. [2,3,4]

1.1.2 The process of creating a PROM

To ensure clinical relevance and scientific soundness, PROMs need to undergo a process of validation after their development. [5]

Indeed, the development and validation of PROMs is guided by international standards. Most notably those set by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN group), the U.S. Food and Drug Administration, and the European Medicines Agency, which provide a comprehensive structure to guarantee these questionnaires are scientifically rigorous and clinically meaningful. [5,6,7]

The COSMIN initiative was created in 2005 with a view to address the variability and sometimes lack of rigor in the creation and assessment of health measurement instruments, including PROMs. [5] The COSMIN consensus recognized the importance of having high-quality, standardized and comparable measurement tools and therefore developed guidelines to validate these instruments. This includes criteria for assessing content and construct validity, structural validity, internal consistency, reliability, measurement error, responsiveness, and interpretability. [5]

The initial phase in developing a PROM involves the generation of relevant domains through a combination of systematic literature reviews, expert input, and, crucially, patient interviews. This ensures that the instrument possesses strong content validity and captures aspects of health that are meaningful from the patient's perspective. Subsequent steps assess the instrument's construct validity, typically by establishing correlations with other PROMs that evaluate similar constructs. [5,8,9]

Reliability of a questionnaire is assessed by examining internal consistency and test-retest stability, while responsiveness measures the PROM's ability to detect clinically important changes over time. Additional steps include defining the minimal clinically important change (MCIC) and, where applicable, performing cross-cultural adaptation and linguistic validation for use in diverse populations. This structured and iterative approach ensures that PROMs are not only psychometrically sound but also relevant to the real-world experiences of patients undergoing treatment for a certain pathology. [8,9]

1.1.3. General vs disease-specific PROMs

PROMs can be broadly categorized into general and disease-specific instruments.

General PROMs assess all-encompassing domains such as overall health perception, pain, physical functioning, and mental well-being. These tools allow for broad comparisons between different populations or diseases but lack sensitivity for small changes in clinical status attributable to a particular disease. [10]

Disease-specific PROMs are tailored to singular conditions and able to better gauge treatment success of a particular disease by detecting smaller but clinically meaningful score changes. [10]

Ideally, in the context of spine surgical care, both types of PROMs should be applied to assess treatment outcome not only on disease specific symptoms but also overall quality of life. [11,12] This stems from the fact that in spine degenerative conditions it is not uncommon to find incongruence between clinical and imaging findings. Indeed, factors beyond mere spinal cord or radicular compression such as psychological well-being, work and socioeconomic status and presence of other musculoskeletal conditions can greatly influence a patient´s functional state. Therefore, a spine disease-specific PROM will more likely miss these factors and overemphasize the impact of spine disease on a patient´s overall clinical state. As such, it should be complemented with a general PROM to give a more holistic health assessment. [11,12]

1.1.4 PROMs in the general medical field

The clinical importance of PROMs was first highlighted in the oncological field. Likely due to its high societal and healthcare impact, oncology was one of the first clinical specialties to streamline the application of these tools in routine clinical care. [13]

Indeed, Barbera et al. performed a retrospective study which showed that follow-up of cancer patients with the Edmonton Symptom Assessment System (ESAS) led to reductions in emergency department visits and hospitalizations.[14] The authors suggest application of the ESAS questionnaire allowed close monitoring of cancer patients´ symptoms and permitted earlier intervention and symptom management. [14]

More recently, Balitsky et al. performed a systematic review and meta-analysis of randomized controlled trials and controlled studies on the impact of PROMs in oncological clinical care and patient experience. [15] The regular application of these questionnaires during follow-up was associated with higher levels of patient satisfaction, increased overall survival, better symptom control and higher quality of life scores. [15]

They also believe patient engagement with PROMs completion leads to earlier symptom detection and management and therefore more timely medical intervention. Additionally, PROMs contribute to the enhancement of patient-physician communication and trust, which are factors linked to better adherence to treatment. [15]

Rheumatology was another discipline where the burden of chronic symptoms, functional limitations, and treatment side effects required more than routine clinical indicators to ascertain the true disease impact on patients´ lives. [16]

The Health Assessment Questionnaire Disability Index (HAQ-DI) was created in the late 1970s by Fries et al. and is one of the earliest and most influential PROMs in clinical medicine. [17]

Initially used in rheumatoid arthritis, the HAQ-DI was designed to systematically capture patients' perspectives on the degree of disability associated with chronic musculoskeletal diseases. The instrument evaluates eight domains of physical functioning that reflect the severity of functional limitations encountered during routine daily activities —dressing, rising, eating, walking, hygiene, reach, grip, and usual activities. The original version also included a ninth domain addressing sexual function, though this is often omitted in clinical and research settings due to sensitivity and variable relevance. [17,18] Its development marked a transformational change in rheumatology by formally incorporating patients' subjective experience of disease into both clinical care and research. Over time, the HAQ-DI

demonstrated strong validity, reliability, and sensitivity to change, and it has since been translated and adapted for use in multiple languages and conditions. [18] The success of the HAQ-DI served as a model for subsequent PROM development across medical specialties, reinforcing the utility of structured patient input in evaluating treatment outcomes. [17]

Consequently, there is some evidence that when properly validated and correctly applied, patient-reported outcome measures can lead to better clinical outcomes and patient satisfaction by facilitating patient follow-up and earlier control of symptoms. [19]

However, as Valderas et al. highlight, further high-quality, methodologically robust studies are needed to provide clear and consistent evidence of these benefits. [19]

This thesis aims to contribute to that critical gap by developing and validating reliable PROM tools specifically tailored for Portuguese patients with degenerative spine surgery, thereby supporting their effective clinical application and ultimately improving patient care in this field.

1.1.5 The challenges lying ahead for PROMs in spine care

The potential benefits of PROMs in obtaining better clinical outcomes have made these tools essential across all medical fields, including degenerative spine disease. Despite the promising results in oncology, the integration of questionnaires into spine care has somewhat lagged behind, with important questions remaining about optimal implementation, validation, and interpretation. [4,20]

Spine surgery presents unique obstacles for PROM implementation. The field is characterized by a high prevalence of subjective, multifactorial symptoms such as pain, discomfort and alterations in sensation, and varying degrees of functional impairment, all of which are influenced not only by the underlying pathology but also by psychosocial and environmental factors. [21]

As Mensah et al. highlight, these subjective dimensions complicate the interpretation of PROMs, as responses are often modulated by patients' individual perceptions, pain tolerance, psychological status, and comorbidities rather than objective disease severity. [21] Moreover, the heterogeneous nature of spine surgery populations - encompassing a wide range of pathologies, surgical techniques, and patient expectations - adds further complexity.

The lack of measurement invariance in commonly used PROMs may also result in significant floor or ceiling effects, limiting their sensitivity across different spine conditions. Finally, the variable indications for surgery — from degenerative disc disease to deformity correction — challenge the standardization of outcome measurement and hinder meaningful comparisons across studies or patient cohorts. Indeed, Mensah et al. argue that without accounting for these

challenges, PROMs risk producing misleading data, reinforcing the need for spine-specific and context-sensitive validation frameworks. [21]

Additionally, many of the widely used spine-specific PROMs — such as the Oswestry Disability Index (ODI) or the Roland-Morris Disability Questionnaire (RMDQ) — were developed decades ago and may lack sensitivity for detecting meaningful clinical change in modern surgical cohorts, particularly in chronic or complex cases. [22] Questions also remain about the minimal clinically important difference (MCID) and / or MCIC across diverse patient groups and surgical procedures. [23] Copay et al. highlight that the MCID is not a fixed value but varies depending on factors such as patient baseline status, the type of surgical intervention, and the specific outcome measure used. [23] This variability complicates the interpretation of PROMs across different clinical contexts, as what constitutes a meaningful change for one patient population or procedure may not apply to another. Moreover, the choice of method to calculate the MCID / MCIC —whether anchor-based or distribution-based—can yield differing reference points, further challenging its broad application. Consequently, establishing context-specific MCIC or MCID values is crucial to accurately assess clinical significance and guide patient-centered care in spine surgery. [23]

As previously mentioned, this thesis aims to explore and address some of these critical gaps. The following chapters of this introduction will investigate the reliability of commonly used instruments in degenerative spine surgery, the determinants and limitations of patient-reported outcomes, and the clinical relevance of PROM score changes in real-world settings.

1.2 PROMs in spine care

Spine care, like general medicine, has shifted from clinician-centered outcomes towards a patient-focused perspective of surgical success.

Spine specific questionnaires have therefore been developed such as the Oswestry Disability Index, Roland-Morris Disability Questionnaire, Neck Disability Index (NDI) and Core Outcome Measures Index (COMI).

1.2.1 Oswestry Disability Index

The Oswestry Disability Index questionnaire was designed in 1980 by Fairbank et al. to measure patient disability associated with back pain. [24]

The ODI assesses 10 key aspects of daily life: pain intensity, ability to perform personal care, lifting, walking, sitting, standing, sleeping, sex life, social life and traveling restrictions. [24]

For this PROM higher scores mean worse clinical outcomes (range: 0 to 100) and each item is scored from 0 to 5. The scores are summed and converted to a percentage, using the formula:

(Total score / Maximum possible score) × 100.

If all items are answered, the maximum score before conversion to a percentage is 50. If some are missing, the maximum is adjusted accordingly.

Based on the total score the original authors divided patients according to five levels of disability: [24]

0–20%: Minimal disability

21–40%: Moderate disability

41–60%: Severe disability

61–80%: Crippling disability

81–100%: Bed-bound or exaggerated symptoms

Although a formal and internationally recognized European Portuguese version of this PROM has not been created, it has been validated to various languages and remains one of the most widely used and reliable questionnaires in lumbar spine care. [25,26]

1.2.2 Roland-Morris Disability Questionnaire

The Roland-Morris Disability Questionnaire is a 24-item "Yes / No" patient-reported outcome measure created to assess disability associated with back pain. [27]

Each question of this questionnaire is worth one point so scores can range from 0 (no disability) to 24 (severe disability).

Chiarotto et al. performed a meta-analysis and systematic review to compare the properties of the RMDQ and ODI and found both PROMs to be equally effective in evaluating low back pain disability in clinical practice. [30]

Although both tools address disability in back pain, the RMDQ seems more suitable for low to-moderate disability. Davidson and Keating compared five questionnaires used for back pain disability, including the ODI and RMDQ, and found the RMDQ more responsive in patients in low disability. [31] Conversely, the ODI performed better in patients with higher disability. Similarly, Roland and Fairbank found the RMDQ more suitable for application in conservative follow-up of patients with low to moderate back disability whereas the ODI provided a more detailed assessment of patients with more severe disability. [22]

1.2.3 Core Outcome Measures Index

The Core Outcome Measures Index was created in 2006 by Mannion et al. as a concise tool to gauge treatment success in patients with degenerative spine disease.[29] This questionnaire focused on 5 domains: pain intensity, function/disability, symptom-specific well-being, quality of life, social and work disability. [29]

Two versions of this PROM were created for patients with degenerative spine disease: the COMI-back for patients with back and / or leg pain and the COMI-neck for assessment of patients with neck pain. [29,32]

For the COMI, higher scores mean worse clinical outcomes (range: 0 to 10). The COMI total score is calculated by averaging the five evaluated domains. Pain is scored from 0 to 10 (using the higher of back / neck or leg / arm pain). The other four domains are scored from 1 to 5 and converted to a 0–10 scale using the formula:

(Score − 1) × 2.5.

After conversion, all five scores are averaged to give the final COMI score.

This questionnaire is efficient and is consistent with other PROMs such as the ODI and Visual Analogue Scale (VAS). [29]

According to Mannion et al., nearly two decades after its development, the COMI has matured into a widely accepted and robust outcome measurement tool in spinal disorder

management. [33] Its brevity and multidimensional focus address key limitations seen in longer, more cumbersome PROMs, facilitating high completion rates and reducing patient burden without sacrificing measurement precision. [33]

Prior to this thesis, the COMI had not been validated to European Portuguese and cross-culturally adapted to the Portuguese population. However, it had been thoroughly validated and cross-culturally adapted to multiple languages and is currently the PROM of choice of the European Spine Society. [33,34,35,36]

The broad application of the COMI in European medical centers also allows for intra and inter-institutional standardization of what constitutes surgical success. [33] Its widespread use and validation support its role as a practical and robust outcome measure in both routine clinical care and spine outcomes research.

1.2.4 Neck Disability Index

The Neck Disability Index was developed in the late 1980s by Dr. Howard Vernon and Dr. Silvano Mior and later validated to European Portuguese by Cruz et al. [28,37]

It was modelled on the ODI and adapted to patients with neck pain. This PROM evaluates the impact of neck pain on daily life by evaluating 10 domains: pain intensity, ability to perform personal care, lifting, reading, headaches, concentration, work, driving, sleeping and recreational activities restrictions.[28]

In similar fashion to the ODI, higher scores in the NDI mean worse clinical outcomes, and the total score is calculated by the same formula.

The authors categorized the raw scores (out of 50) into five levels of disability:[28]

0–4: No disability

5–14: Mild disability

15–24: Moderate disability

25–34: Severe disability

35–50: Complete disability

The NDI has become one of the most frequently used and accepted tools for assessing neck pain–related disability due to its simplicity, ease of administration, and clinical relevance.

MacDermid et al. performed a systematic review that confirmed the robust measurement properties of the NDI across diverse settings and patient populations.[38] The review synthesized findings from multiple validation studies and provided strong evidence for the NDI's reliability, validity, and responsiveness. Internal consistency was repeatedly reported as good (with

Cronbach's alpha values typically exceeding 0.80), indicating that the items reliably measure a common construct.[38] Test-retest reliability was also strong, demonstrating that the NDI yields stable scores in clinically stable patients over time.[38]

Construct validity was supported through correlations with other measures of pain and function, and the instrument showed good discriminative ability between groups with differing levels of disability.[38] Responsiveness—the ability to detect clinically meaningful changes over time—was confirmed in several longitudinal studies, although the review highlighted that responsiveness could vary depending on the intervention and the specific population studied.[38]

Despite its strengths, MacDermid et al. also discussed limitations in the NDI's measurement framework.[38] The NDI was designed to be unidimensional but factor analysis in different populations has shown that the items sometimes group into separate domains. This implies that the NDI may be capturing multiple dimensions of the disability experience — possibly reflecting both physical and psychological components of neck pain — rather than a single, unified construct.

These findings underscore the importance of contextual and methodological considerations when applying the NDI in research or clinical practice.

Although the NDI might not be a purely single-score (unidimensional) measure, this PROM remains highly relevant in neck pain assessment. It is practical, quick to administer, and provides a validated quantitative measure of functional status. The review by MacDermid et al. solidifies its position as a reliable PROM, while also highlighting areas for refinement and encouraging further research into population-specific and culturally adapted versions. [38]

1.2.5 General health-related PROMs in spine care

To mitigate the weaknesses of disease-specific PROMs and broaden the scope of health assessment, spine questionnaires are usually complemented in clinical practice and research with general health-related PROMs such as the EuroQol – 5D (EQ-5D) and the Short Form 36 Health Survey Questionnaire (SF-36). [39,40] These tools assess patient general quality of life and the impact of spine pathology on overall health.

The EQ-5D evaluates mobility, ability to perform self-care and routine daily activities, pain / well-being and depression / anxiety and has been validated to European Portuguese. [41]

For this PROM higher scores translate into better clinical results (maximum score: 1).

The SF-36 comprehensively reviews physical and mental health by evaluating eight domains: limitations in physical activities because of health problems and in social activities because of physical or emotional problems, limitations in usual role activities because of physical health

problems, bodily pain, general mental health (psychological distress and well-being), limitations in usual role activities because of emotional problems, vitality (energy and fatigue); and general health perceptions. [40]

The SF-36 does not have a unique "global score." Instead, each domain has a score ranging from 0 to 100, with higher scores indicating better health.

This tool has been validated to European Portuguese and is highly sensitive to alterations in clinical status, however it is more complex and takes longer to fulfil than the EQ-5D. [41] Therefore, the EQ-5D, due to its brevity and ease of use, is better suited for larger studies while the SF-36 can provide a more patient-specific detailed clinical assessment. [43]

McDonough et al. compared the EQ-5D and SF-36 in patients with spinal conditions and found that the instruments capture different aspects of health, reinforcing their complementary use depending on the goals of the study or clinical evaluation. [44]

However, and in contrast to the EQ-5D, there is no Portuguese MCID or MCIC to assess clinical improvement for the SF-36.

Thus, while both PROMs play complementary roles in spine care assessment, ongoing research to establish clinically meaningful change thresholds for the SF-36 in Portuguese populations would enhance its utility in both clinical practice and research settings.

1.2.6 MCIC and MCID in spine care

The minimal clinically important change refers to the smallest change in PROM score that is perceived by the patient as representing meaningful improvement in their clinical status. [45] Albeit often used interchangeably, there are differences between the MCIC and the minimum clinically importance difference. Whilst the MCIC translates a difference in single patient status across a longitudinal clinical evaluation, the MCID refers to the smallest change in PROM score that is clinically significant when comparing groups of patients. [46]

Various methods have been created to calculate the MCID and MCIC. The following overview synthesizes established methodologies and recommendations from seminal works by Jaeschke et al. (1989), Ostelo et al. (2008), and Revicki et al. (2008), which have shaped current consensus on determining meaningful change in patient-reported outcomes. [45,46,47]

First, anchor-based methods rely on an external reference—or "anchor"—such as a patient-reported global rating of change (e.g., "Much improved," "Slightly better"). Patients are grouped based on their response to the anchor, and the average change in PROM scores is calculated for each group. The smallest statistically significant difference between groups is identified as the MCID. [45]

A receiver operating characteristic (ROC) curve analysis is often applied within this method and determines the optimal cutoff score based on the point of highest combined sensitivity and specificity. When focused on within-patient change over time, anchor-based methods can also be used to estimate the MCIC.

Second, distribution-based methods estimate the MCID by examining the statistical properties of PROM score distributions in the study sample. [47] Common metrics include effect size (ES), standard deviation (SD), and standard error of measurement (SEM). These approaches do not incorporate patient perception but provide objective thresholds that can complement anchor-based results. [47]

Third, regression-based methods combine aspects of both anchor-based and statistical modeling. Here, the PROM change score is the dependent variable, and the anchor—often dichotomized as "improved" vs "not improved"—is the independent variable. This method allows for greater statistical rigor and flexibility but is also more complex and potentially less intuitive. [47]

Anchor-based methods are more straightforward and simpler but less precise, whilst regression-based methods offer greater flexibility and statistical rigor at the cost of complexity and interpretability. [47] The choice of method should be decided according to the study design, sample size, and the intended application of the MCIC estimate.

It is of utmost importance to establish an MCIC or MCID in spine PROMs to improve the interpretability of these tools in clinical care and better understand what constitutes meaningful treatment outcome from a patient-centered perspective.

1.3 Limitations and pitfalls of PROMs

1.3.1 The subjectiveness of PROMs

Patient-reported outcome measures have limitations that should be addressed in clinical care to strengthen their accurate interpretation.

Firstly, PROMs are inherently subjective because patient well-being or disease is not a fixed construct. Sprangers and Schwartz described the concept of "*response shift*" which represents a change in patient´s perception of well-being as a disease process develops. [48] Therefore, a chronic condition might alter patients' perception of pain or disability due to symptom under- or overestimation as disease adaptation takes place. This subjectivity can limit longitudinal interpretation of PROMs. [48]

Indeed, Fayers and Machin acknowledge this inherent natural characteristic of PROMs that should be accounted for when applying questionnaires in clinical practice and research.[49] They recommend the use of reliable and valid tools, repeated measurements at different timepoints to adjust for intra-patient variability and establishment of a minimal clinical important difference for each questionnaire to improve the interpretability of PROMs.[49]

Accounting for this subjectivity is essential to improve the clinical utility of PROMs in decision-making and treatment evaluation, particularly in spine care where perceived disability often diverges from radiological findings.

1.3.2 Measurement Invariance

Another important limitation of PROMs relates to their measurement invariance—the extent to which the questionnaire measures the same underlying construct across different subgroups of patients: different genders, ages, cultural or linguistic backgrounds. Without establishing measurement invariance, comparisons of PROM scores across groups may be misleading, as differences could reflect measurement bias rather than true differences in the construct of interest.[50] This is especially important for PROMs that have been translated and adapted to different cultures, because small differences in meaning can change how questions work and affect the comparability of scores.[51] As van de Vijver and Tanzer explain, such subtle differences can lead to inequivalence in measurement, with scores no longer reflecting the same psychological or health construct across populations. Consequently, direct comparisons between groups—such as between countries, languages, or demographic strata—may lack validity and lead to erroneous clinical or research conclusions. [51]

Testing measurement invariance through techniques like multi-group confirmatory factor analysis is therefore essential to confirm that PROMs provide valid and comparable data across diverse patient populations.[52] According to Putnick and Bornstein this testing happens in stages: first, it checks whether the overall structure of the questionnaire is similar across groups; then, it verifies if each question relates to the main concept in the same way for everyone; and finally, it confirms that people with the same level of the trait being measured would get similar scores, no matter which group they belong to. Each step contributes to confirming that the PROM measures the same construct across different populations. [52]

Failure to account for invariance may result in biased clinical interpretation and research conclusions, limiting the generalizability and equity of PROM use in spine care. Even so, all PROMs in spine care mentioned in this research have been validated and cross-culturally adapted without a specific emphasis on measurement invariance. Reasons may include statistical complexity in performing multi-group confirmatory factor analysis, population homogeneity within most cultural validation contexts and absence of measurement invariance as an explicit requirement in COSMIN guidelines.

1.3.3 Floor and ceiling effects

Floor and ceiling effects can also hamper the sensitivity and interpretation of PROMs by limiting understanding of relevant changes to patients´ condition.

Floor effects occur when patients already score at the worst end of a PROM, making it difficult to detect further deterioration in their condition after intervention. Ceiling effects occur when patients are at the lower end of disability and are unlikely to signal improvement in their clinical condition with treatment. [26,53]

Measures to handle these effects include creating PROMs with a wide range of responses, namely increasing the values in a pain score scale (e.g., from 5-point to 11-point scales) or adding more answers to questions addressing the different domains in the questionnaire. [8] The use of both disease-specific and general health questionnaires can also improve the sensitivity to detect minor changes in clinical status and well-being whereas longitudinal assessments with repeated measurements also allow for detection of subtle changes in clinical status. [26,53]

1.3.4 Logistical challenges in the application of PROMs

Another limitation of PROMs in clinical practice relates to the administrative and time-consuming nature of routine questionnaire application.

Physicians and administrative staff may find it challenging to implement and fit questionnaires into a busy schedule as it can cause delays to workflow. [54] Patients may also perceive some PROMs to be too lengthy and burdensome, especially if they are not well-informed about the potential benefits to patient care and research of PROMs. [54,55]

To overcome this limitation, Stover et al. propose that patients and physicians be fully informed about the importance of these tools to overall healthcare and clinical research. [56]

Improving patient and physician awareness, streamlining administration processes, and integrating PROMs into digital workflows are essential measures to help ensure their sustained and meaningful use in spine care.

1.3.5 Electronic PROMs, health illiteracy and socioeconomic impact

Electronic PROMs (ePROMs) have recently started to be used in clinical practice to reduce some of the administrative workload and streamline data collection. However, digital illiteracy and reduced access to digital platforms in certain populations, difficulty integrating ePROMs with electronic health records and data security are some of the challenges of these tools. [57]

Indeed, health and educational illiteracy is another possible hindrance to adequate use of both types of PROMs. Patients may not understand medical terms or adequately interpret the questions; therefore, patient education level needs to be accounted for when applying these tools. [58]

Tuinenburg et al. performed a review on the comprehensibility of 157 PROMs applied in healthcare to the Dutch population. [59] They used the Pharos Checklist for Questionnaires in Healthcare (PCQH) to understand whether comprehensibility was obtained in 8 different domains: language level, number of questions and answer options, presence of instructions, concrete questions, medical terms, questions in active voice and statements. Overall optimal comprehensibility across all domains was seen in only 11% of the questionnaires, which highlights the need for further optimization of PROMs interpretability. [59]

To help patients better interpret the questionnaires and overcome some of these limitations PROMs should be validated and cross-culturally adapted to the respective language and population.[60]

However, in populations with lower educational levels or immigrant populations that lack full proficiency of the native language those limitations may remain to be fully addressed.

Konopka et al. performed a retrospective review of 16119 cases of patients submitted to knee or hip joint arthroplasty in the United States of America and found patients of non-English speaking minorities, from lower-income neighborhoods and those less educated to not only participate less in PROMs completion but also to report worse clinical outcomes pre- and post-operatively. [61] This study strengthens the importance of validating and cross-culturally adapting PROMs to every population and to creating healthcare support systems that engage and assist disadvantaged populations with the fulfilment of these questionnaires.[61]

These findings are also highly relevant to spine care, where PROMs are central to patient-centered assessment but must be equitably accessible and interpretable across all patient demographics.

1.3.6 Recall bias

Another pitfall of PROMs is recall bias. Recall bias refers to the inability of patients to accurately remember past health status and level of disability that can lead to under-or overestimation of current symptoms. [62]

Moreover, recall bias can impact the accurate assessment of post-operative PROM scores. Recall bias can also affect anchor-based methods as they depend on patients' recall of their prior clinical status, which can introduce a risk of misinterpreting the minimal clinically important difference if patients fail to accurately remember their preoperative condition. [63]

Macchiarola et al. analyzed children with different knee pathologies who completed an evaluation of the same questionnaire before and twelve months after surgery. [64] They found a poor correlation between assessments that led to an underestimation of their pre surgical health status in 82% of the cases. [64]

Similarly, Gotlin et al. demonstrated poor agreement between PROMs collected retrospectively and prospectively. This led to overestimation of surgical outcome in adult patients that underwent rotator cuff surgery when relying on retrospective recall. [65]

Recall bias has also impacted patient-reported outcome measures in spine care. Pellisé et al. compared the results of 3 PROMs completed before posterior lumbar instrumentation surgery and postoperatively at a mean of 37.5 months after surgery. [66] They found that retrospectively assessing questionnaires led to inaccurate PROMs recall and patient overestimation of surgical improvement. [66]

Rodrigues et al. analyzed various questionnaires in patients with lumbar and cervical degenerative spine disease. [62] For lumbar conditions, they assessed the COMI-back, the Oswestry Disability Index, and a Numeric Rating Scale (NRS) for back and leg pain. For

cervical conditions, the authors used the COMI-neck, the Neck Disability Index, and the NRS to evaluate cervical and arm pain. Questionnaires collected 12 months after surgery were compared to those collected before surgery and they found patients overestimated the degree of disability previous to surgery, particularly patients with lumbar disease and those with greater clinical improvement. [62]

Measures to address recall bias were proposed in an influential paper by Karen Raphael in 1987 and included prospective data collection, avoiding reliance on long-term memory and having strong, validated and standardized questionnaires. [67]

Given the demonstrated impact of recall bias on both under- and overestimation of symptoms and treatment outcomes, it is essential that clinicians and researchers account for this limitation when utilizing PROMs. Prospective data collection methods, minimizing reliance on patients' long-term memory, and employing validated, standardized questionnaires can help mitigate recall bias and enhance the accuracy of patient-reported data.

1.3.7 Response biases

When completing PROMs, patients may be conditioned by knowing their physician will read and analyze their answers. This may stem from the usual power asymmetrical doctor-patient relationship. [68] Zini and Banfi performed a literature review highlighting that the awareness of clinician oversight can lead patients to alter their responses consciously or unconsciously. [69] There are concerns patients may report better outcomes not to disappoint the surgeon and justify a previous surgical intervention, particularly if they have a good relationship with their physician. Patients may also downplay symptoms to avoid being considered complainers or in order to avoid more interventions. Conversely, mismatched expectations before surgery where patients expected greater clinical changes may lead patients to overreport symptoms albeit an objective improvement in clinical status occurred. This may distort results negatively and hamper interpretation of postoperative scores.

However, recent evidence challenges this assumption. In 2024 Haverman et al. conducted secondary analyses of three studies assessing PROMs in adult and pediatric care. [70] They compared two groups: in the intervention group patients were informed that the results would be shown to their physicians and in the control group patients were told that their clinicians would not see their responses. They found no significant differences between PROM scores in the two groups, which did not support the view that patients bias their responses if they know their physicians will have access to their answers. [70]

Nonetheless, in a bid to strengthen PROM interpretability and address these potential biases measures can be implemented during the application of clinical measurement tools. These may include blinding the clinician to PROM responses and assuring patients of this anonymity. Additionally, PROMs might be complemented with patient-directed instructions emphasizing the importance of honest, independent feedback for improving care quality.

1.4 Towards valid and reliable tools to assess surgical success in degenerative spine surgery

The previous chapters highlighted the potential relevance of different PROMs in clinical and spine care by reinforcing the patient-physician relationship and providing a more objective assessment of treatment outcomes. However, there are multiple challenges that still limit the full application and integration of these tools in spine care.

Medicine is a fast-evolving field, where machine learning and artificial intelligence are emerging as important tools and may help overcome some of the challenges previously described. These emerging technologies have the potential to facilitate data acquisition and reduce the administrative burden on health professionals and patients alike. Therefore, these technologies may enable physicians to focus more on the human side of medicine by fostering closer patient-doctor relationships. Hence, a shift towards the widespread application of ePROMs with the aid of these technologies and reinforcement of digital literacy can help patients better understand the value of these tools in improving their care.

In the following chapters we define the aims of this thesis and discuss the findings of the research that sought to create valid and reliable tools to assess surgical success in degenerative spine surgery.

2. AIMS

2.1 Thesis timeline

The recall bias study was conducted first to address a key methodological concern in PROM-based outcome research: the accuracy of retrospectively recalled preoperative data. This was particularly relevant because we used an existing MCID threshold (an improvement of 2.6 points on the COMI scale defined for another European population [71]) to classify clinical improvement, relying on patients' recalled baseline status. While this issue does not directly impact PROM validation, it was important to establish early whether retrospective assessments could be trusted in our broader research context. Doing so helped clarify the limitations of retrospective designs and informed the methodological rigor of subsequent studies.

2.2 Thesis aims

The research that was undertaken to produce this PhD thesis had three general aims:

- To evaluate the impact of recall bias of preoperative clinical status on the postoperative scores of patient-reported outcome measures in patients submitted to surgery due to degenerative disease of the lumbar and cervical spine.

- To cross-culturally adapt and validate the Core Outcome Measures Index-back to European Portuguese and establish the minimal clinically important change for this version in patients with degenerative disease of the lumbar spine.

- To cross-culturally adapt and validate the Core Outcome Measures Index-neck to European Portuguese and establish the minimal clinically important change for this version in patients with degenerative disease of the cervical spine.

Three studies were conducted to achieve these aims:

1. A randomized controlled trial between two groups with similar clinical characteristics was undertaken. It aimed to ascertain whether providing the preoperative PROM answers to patients that underwent surgery due to degenerative spine disease led to different postoperative PROM scores. Postoperative scores in the questionnaires were compared to patients not given access to their preoperative scores (control group). Therefore, this study sought to mitigate recall bias by presenting to patients in the intervention group their preoperative answers when completing postoperative assessments.

2. This study followed published guidelines to validate and cross culturally adapt the COMI questionnaire into European Portuguese and the Portuguese population with degenerative disease of the lumbar spine. Moreover, the minimal clinically important change was obtained for this PROM using an anchor-based method.

3. This study followed published guidelines to validate and cross culturally adapt the COMI questionnaire into European Portuguese and the Portuguese population with degenerative disease of the cervical spine. Additionally, an MCIC was obtained for this PROM using an anchor-based method.

Together, these studies aimed to enhance the accuracy, applicability, and clinical utility of PROMs in spine surgery outcome assessment.

3. PUBLICATIONS

3.1 PAPER A

Are the results of patient-reported outcome measures after spine surgery influenced by recall of preoperative scores? A randomized controlled trial

3.2 PAPER B

Cross-cultural adaptation, validation and establishment of the minimal clinically important change score of the European Portuguese core outcome measures index in patients with lumbar degenerative spine disease

3.3 PAPER C

Cross-cultural adaptation, validation, and establishment of the minimal clinically important change score of the European Portuguese Core Outcome Measures Index in patients with cervical degenerative spine disease

3.1 PAPER A

ARTICLE IN PRESS

ELSEVIER

The SPINE JOURNAL

The Spine Journal 000 (2022) 1–10

Clinical Study

Are the results of patient reported outcome measures after spine surgery influenced by recall of preoperative scores? – a randomized controlled trial

Pedro Valente Aguiar, MD[a,b,*], Pedro Santos Silva, MD[a,b], Rui Vaz, MD, PhD[a,b,c], Paulo Pereira, MD, PhD[a,b,c]

[a] *Neurosurgery Department, Centro Hospitalar Universitário São João, Porto, Portugal*
[b] *Faculty of Medicine, Porto University, Porto, Portugal*
[c] *Neurosciences Centre, CUF Porto, Porto, Portugal*

Received 8 June 2022; revised 13 October 2022; accepted 8 November 2022

Abstract

BACKGROUND CONTEXT: Patient reported outcome measures (PROMs) are of utmost importance to clinical practice as they permit a patient-focused evaluation of surgical outcomes. However, recall bias can limit an adequate interpretation of PROMs.
PURPOSE: To assess the impact of recall bias of preoperative status on postoperative PROMs of patients submitted to surgery due to degenerative spine disease.
STUDY DESIGN / SETTING: Randomized controlled trial in a tertiary care neurosurgical unit in Portugal
PATIENT SAMPLE: All patients submitted to surgery at our institution from January 2019 to April 2020 due to degenerative lumbar or cervical spine disease with valid PROMs questionnaires were enrolled, and 2 computer generated randomized groups were created.
OUTCOME MEASURES: The study's primary endpoint was the median postoperative Core Outcome Measure Index (COMI) score.
METHODS: The intervention group was sent postoperative questionnaires including preoperative answers, while patients in the control group were sent the same PROMs without the preoperative answers.
RESULTS: Randomization was applied to 236 patients (118 for each group) and valid results were obtained for 147 patients (81 lumbar, 44 from the intervention group; and 66 cervical, 29 from the intervention group), from which 88 (60%) were females, with a median age of 58 years.

Both groups shared similar baseline clinical characteristics and preoperative scores. Median postoperative COMI scores and interquartile ranges (IQR) were 4.20 (IQR: 2.30–6.00) and 5.45 (IQR: 3.75–7.40) for the intervention and control groups, respectively (Wilcoxon, p=.02). This difference was reached mainly due to cervical spine patients as median postoperative COMI score was 3.95 (IQR: 2.20–5.32) in the intervention group and 5.1 (IQR: 4.0–8.4) in the control group (Wilcoxon, p=.01). No significant difference was reached for lumbar patients.
CONCLUSIONS: Better PROMs scores were obtained for degenerative cervical spine patients to whom the preoperative results were provided. Therefore, providing preoperative scores to patients upon postoperative PROMs fulfilment might influence postoperative results. Further research is necessary to increase the reliability of PROMs in clinical practice.

Keywords: Degenerative spine disease; Memory bias; PROMs; MCID; COMI cervical; RCT

FDA device/drug status: Not applicable.
Author disclosures: ***PVA:*** Nothing to disclose. ***PSS:*** Nothing to disclose. ***RV:*** Nothing to disclose. ***PP:*** Nothing to disclose.
Trial registration trial is registered at clinicaltrials.gov ID: CHSJ1989

*Corresponding author. Centro Hospitalar Universitário São João, Alameda Prof Hernâni Monteiro, 4200-319 Porto
E-mail address: pedrovalenteaguiar@protonmail.com (P. Valente Aguiar).

https://doi.org/10.1016/j.spinee.2022.11.007
1529-9430/

Introduction

Spine surgery aims to improve patient well-being by enhancing neurological status, alleviating pain, and improving function. In recent decades increased emphasis has been put on assessing patients' self-perceived improvement following both medical and surgical care through patient-reported outcome measures (PROMs) [1,2].

Indeed, Basch et al. showed in a randomized controlled trial (RCT) that the use of PROMs could lead to better clinical outcomes and quality of life in cancer patients who regularly reported their symptoms in comparison to patients that were not frequently asked about their health status [3].

Nevertheless, PROMs have certain limitations that ought to be addressed to further increase clinical utility. Those limitations range from technical and time constraint difficulties when applying PROMs in daily clinical practice, to inadequate health practitioner-patient discussion regarding their significance [4–6].

Of particular interest to patients with degenerative spine disease, recall bias has been shown to impact score results, therefore rendering retrospective assessment of preoperative patient reported outcomes less reliable [7,8]. Aleen et al. reported inadequate patient recall of preoperative status in individuals submitted to lumbar decompression and fusion with over 40% of patients not properly recalling their main preoperative symptom [8]. Aleen and colleagues also found degenerative cervical spine patients to have poor recall of their preoperative symptoms, with over 30% reporting a distinct main preoperative symptom when filling PROMs a year after surgical intervention [7]. Additionally, Rodrigues et al. found patients submitted to degenerative lumbar surgery to have poor recollection of their preoperative status leading to worse recalled scores in comparison to actual preoperative scores [9].

Based on this perception and findings, we hypothesized that patient recall bias might also lead to altered postoperative scores, as failure to recollect preoperative status may underestimate or overestimate postsurgical improvement in health status.

In this RCT, we aimed to evaluate the impact of recall bias on postoperative PROMs by providing to the intervention group their preoperative scores upon collecting their postoperative PROMs (minimum 1 year after surgery), and compare these to postoperative PROMs in patients not informed about their preoperative results in the same timeframe.

Methods

Trial design, participants and setting

This randomized controlled trial was approved by the ethics committee of our institution and followed the CONSORT statement for RCTs and the CONSORT PRO Extension for reporting of patient-reported outcomes in randomized trials [10,11].

All adult patients with valid preoperative PROMs scores submitted to spine surgery due to degenerative pathology from January 2019 to April 2020 at our center (spine unit of a neurosurgical department in a Portuguese university hospital) were included in the study and received PROMs questionnaires with a letter of consent.

Questionnaires were sent by post to all patients and if PROMs were not returned within 4 weeks, patients were contacted by phone to confirm whether they had received the mail. PROMs were resent if patients mentioned they had failed to receive the questionnaires.

Patients without valid preoperative questionnaires were excluded from the study while patients were lost to follow-up if mail questionnaires were not completed 2 months after PROMs remittance or if they had submitted incomplete or invalid answers.

Interventions

To study the impact of recall bias on the scores of postoperative PROMs the patients of the intervention/recall group received postoperative questionnaires that included the preoperative answers for every individual question, whilst the control/no-recall group answered postoperative questionnaires without the preoperative responses.

Patients' clinical characteristics including age, gender, body mass index (BMI), education level, main clinical presentation, surgery date and surgical outcome were retrieved both from the clinical registries at our institution and the EUROSPINE Spine Tango registry.

In preoperative and postoperative evaluation, to assess surgical outcome and disability, lumbar spine patients filled the back Core Outcome Measures Index (COMI) and Oswestry Disability Index (ODI) questionnaires, while patients with cervical degenerative spine disease completed the neck surgical COMI and the Neck Disability Index (NDI) forms. For these questionnaires higher scores mean worse clinical outcomes (COMI range: 0 to 10) (NDI and ODI range: 0 to 100). The EuroQol Five Dimension Questionnaire (EQ-5D) was used to evaluate quality of life for both lumbar and cervical patients, and conversely for this PROM higher scores translate into better clinical results (maximum score: 1).

All these questionnaires have proven their reliability and validity in degenerative spine patients of different populations and have been thoroughly used to assess patient outcomes [12–15].

In order to evaluate meaningful clinical improvement, we considered the minimal clinical important difference (MCID) for COMI as 2.6 based on the work by Mannion et al in a European centre [12].

The considered MCID was 14.9 for the ODI and 17.3% for the NDI, and respectively 0.24 and 0.49 for cervical and lumbar patients for the EQ-5D, based on the works by Parker and colleagues [13,14].

To rule out discrepancies with regards to patient outcomes between the intervention and control groups, clinical scores following Odom's criteria were defined based on an evaluation by the treating surgeon that attributed 4,3,2 or 1 point to excellent, good, fair, and poor clinical outcomes respectively [16].

A similar scoring system was employed to calculate median scores from the answers of the postoperative COMI questionnaires with regards to satisfaction and outcome perceived by patients, with 5 points given to the highest level of patient satisfaction and 1 to the lowest.

The first author and main investigator was responsible for PROMs remittance and evaluation but was not part of the surgical team nor followed the patients in the outpatient clinic.

Outcomes

The primary outcome of the study was the postoperative COMI score, while the secondary outcomes were the postoperative NDI or ODI and EQ-5D scores. Additionally, subgroup analysis was performed to evaluate the differences in these postoperative scores for lumbar and cervical patients. Item analysis for COMI questionnaire was performed to establish whether recall bias had higher impact on specific questions.

Sample size

Effect size was obtained according to the quotient of minimal detectable change (MDC) for COMI and the standard deviation (SD) of the minimal clinical important difference (MCID).

In the absence of a European Portuguese MDC, we used the value of 1.7 set by Damasceno et al. in a Brazilian – Portuguese population while the MCID standard deviation (2.7–2.9) was based on the work by Mannion et al. in a European centre [17,18].

Sample size was calculated using G*Power 3.1.9.4 for a 2-sided Wilcoxon-Mann-Whitney test, with significance level of 0.05 and a power of 0.95 to a final value of 150 patients. The enrollment accounted for a 30 to 40% expected loss of patients to follow-up during study completion.

Randomization

A simple randomization procedure was performed using a random number generator (Microsoft Excel) to 2 independent lists corresponding to cervical and lumbar patients; each list was sorted in ascending order according to an assigned random number and split in half. Subsequently 2 groups of equal size for each spine segment were generated, corresponding to the control and intervention groups.

The generation of the randomization which included patient enrollment and assignment to trial groups was done by a third-party whereas the first author conducted the implementation of the randomization.

Due to the nature of the study, blinding was deemed unnecessary and was not undertaken.

Statistical methods

R software (R Foundation for Statistical Computing, Vienna, Austria) version 4.0.3 was used for data analysis. The code and additional analysis are presented in a supplemental file.

Kruskal-Wallis or Wilcoxon-Mann-Whitney tests were used for median comparison between independent samples and for ordinal data, and Fisher's test was used for associations between categorical variables.

Classical test theory was used for item analysis. COMI items related to pain (2a. and 2b.) were transposed to a 0 to 5 scale, then standardized Cronbach's alpha was used for reliability comparison between recall and no-recall groups. Individual item analysis was performed with alpha-if-item-omitted and item-to-total correlation values.

Results

Participant flow and recruitment

From January 2019 to April 2020, 424 adult patients were submitted to degenerative spine surgery at our department, 188 of which were excluded due to invalid or absent preoperative scores.

A total of 96 cervical and 140 lumbar patients were enrolled in this study and valid postoperative scores were obtained for 147 patients, 59 (40.1%) males and 88 (59.9%) females with a median age of 58 years. The mean time between surgery and questionnaire fulfilment was 17 months.

Due to absence of postoperative questionnaires or invalid answers, 89 patients were lost to follow-up. (Fig. 1 shows the patients' flow diagram).

Outcomes and estimation

Seventy-four answers were retrieved from the control group and 73 from the intervention group, with both displaying similar demographic and clinical characteristics, including gender, body mass index (BMI), smoking status, rates of fusion and minimally invasive surgery, educational status, and main clinical presentation. The difference in median age between the no-recall and recall subgroups (56 vs 62 years) was considered clinically irrelevant. (Table 1 displays patient characteristics)

Both lumbar and cervical patients lost to follow-up shared similar clinical characteristics and preoperative PROMs scores. (supplementary file)

The preoperative status was similar between groups, as there were no significant differences in preoperative PROMs, namely COMI, ODI/NDI and EQ-5D (Table 1).

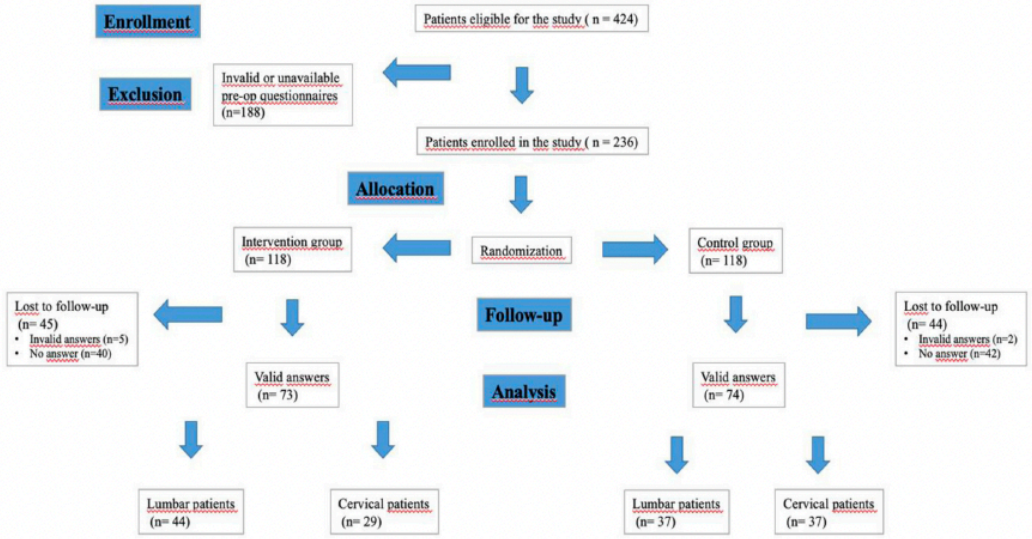

Fig. 1. Flow diagram.

Outcomes for surgical results measured through surgeon (Odom's criteria) (Fig. 2) and patient perceptions (answers to COMI's outcome and satisfaction questions) showed no statistical difference between groups (Table 1).

Postoperative PROM scores

Regarding the postoperative COMI, the primary endpoint, the median scores, and interquartile ranges (IQR) were 5.45 (IQR: 3.75−7.40) for the control group and 4.20 (IQR: 2.30−6.00) in the recall group (Wilcoxon, p=.023), and although statistically significant, the difference did not achieve the effect size defined for the power of the study. This difference between groups was also observed with functional scores, the NDI or ODI, with a median of 36.00 (IQR: 24.00 − 54.00) for the control group and of 28.45 (IQR: 14.00−39.50) for the recall group (Wilcoxon, p=.006). Quality of life as measured by EQ-5D was also different between groups: median of 0.586 (IQR: 0.452 − 0.691) for the control group and 0.689 (IQR: 0.516 − 0.813) for the recall group (Wilcoxon, p=.017) (Fig. 3).

In a subgroup analysis by spine segment, it becomes clear that these differences in postoperative scores can be mainly attributed to patients with cervical disease (Fig. 4).

Although better clinical scores were marginally observed in the intervention group of lumbar patients, they did not reach statistical significance. (supplementary file) For this reason, we decided to analyze separately the subgroup of patients submitted to cervical surgery.

Cervical surgery patients

Sixty-six patients with cervical pathology were included in the study, 37 (56.0%) from the no-recall group and 29 (44,0%) from the recall group. The mean time between surgery and questionnaire fulfilment was 16.5 months.

As for overall patients, the preoperative status of cervical patients was similar between groups, as there were not significant differences in preoperative PROMs (Table 2).

Regarding postoperative PROMs the differences between groups were statistically significant for the 3 scores (Fig. 5).

The median of postoperative COMI for the no-recall group was 5.10 (IQR: 4.00−8.40) and 3.95 (IQR: 2.20 −5.33) for the recall group (Wilcoxon, p=.011) while for postoperative NDI the median was 44.00 (IQR: 31.00 −54.00) in the no-recall group and 24.40 (IQR: 11.00 −36.70) in the recall group (Wilcoxon, p<.001). The postoperative difference between groups for EQ-5D in cervical patients was 0.173 (medians of 0.516 vs 0.689, Wilcoxon p=.015), a change in favor of the recall group.

Regarding median improvement in the PROMs, the cervical control group had the worst results (Table 3) as both lumbar (no-recall and recall) and cervical recall groups reported clinically meaningful results whilst the no-recall cervical group had median improvements with surgery below MCID for all questionnaires.

Regarding reliability of the COMI questionnaire, measured with standardized Cronbach's alpha, the items responses from both groups showed good internal consistency (α=0.906 for no-recall and α=0.869 for recall group).

Table 1
Overall patients' characteristics

	no recall (N=74)	recall (N=73)	p value
Gender			.737*
Female	43 (58.1%)	45 (61.6%)	
Male	31 (41.9%)	28 (38.4%)	
Age			.035†
Median (Q1,Q3)	56 (49.000, 63.000)	62.000 (52.000, 69.000)	
Segment			
Cervical	37 (50.0%)	29 (39.7%)	
Lumbar	37 (50.0%)	44 (60.3%)	
Main clinical presentation			.415*
cervical radiculopathy	13 (17.6%)	10 (13.7%)	
lumbar radiculopathy	24 (32.4%)	23 (31.5%)	
myelopathy	24 (32.4%)	19 (26.0%)	
neurogenic claudication	13 (17.6%)	21 (28.8%)	
BMI			.822†
Median (Q1,Q3)	27.250 (25.3, 29.940)	26.915 (24.8, 29,678)	
Smoker			.265*
yes	15 (20.5%)	9 (12.7%)	
no	58 (79.5%)	62 (87.3%)	
Education			1.000*
≤4 years	26 (48.1%)	24 (48.0%)	
>4 years	28 (51.9%)	26 (52.0%)	
Preoperative COMI			.557†
Median (Q1, Q3)	8.1 (6.5, 8.8)	7.8 (6.7, 8.8)	
Preoperative ODI/NDI			.151†
Median (Q1, Q3)	53.5 (40, 64.3)	48 (37.9, 58.5)	
Preoperative EQ5D			.153†
Median (Q1, Q3)	0.516 (-0.013, 0.587)	0.516 (0.088, 0.620)	
Postoperative COMI			.023†
Median (Q1, Q3)	5.5 (3.8, 7.4)	4.2 (2.3, 6)	
Postoperative ODI/NDI			.006†
Median (Q1, Q3)	36(24, 54)	28.5 (14, 39.5)	
Postoperative EQ5D			.017†
Median (Q1, Q3)	0.586 (0.452, 0.691)	0.689 (0.516, 0.813)	
Satisfaction (COMI)			.874†
Median (Q1, Q3)	5 (5, 5)	5 (5, 5)	
Outcome (COMI)			.116†
Median (Q1, Q3)	4 (3, 5)	4 (3, 5)	
Odom			.638†
poor	11 (14.9%)	8 (11.0%)	
fair	21 (28.4%)	25 (34.2%)	
good	32 (43.2%)	25 (34.2%)	
excellent	10 (13.5%)	15 (20.5%)	
COMI improvement			.047†
Median (Q1, Q3)	1.76 (0.4, 4.4)	3.000 (1, 4.8)	
ODI/NDI improvement			.263†
Median (Q1, Q3)	11 (1.6, 24.7)	16.3 (5.7, 29.6)	
EQ5D improvement			.407†
Median (Q1, Q3)	0.104 (-0.034, 0.467)	0.188 (0.000, 0.448)	

* Fisher Exact Test for Count Data.
† Kruskal-Wallis rank sum test.

The item analysis (supplementary file) revealed that for every item, patients with recall scored lower. In the no-recall group the questions 2b (numeric rating scale [NRS] arm/shoulder pain) and 4 (symptom specific well-being) had more negative impact on Cronbach's alpha (alpha-if-item-omitted of 0.903 and 0.905 respectively). Moreover, an arm/shoulder NRS >=7 was reported by 50% of no-recall but only 25.9% of recall patients.

Discussion

Patient reported outcomes are fundamental in health care and offer additional advantages to the traditional clinical and radiological findings used to define clinical outcome. However, among the various issues that should be optimized to further increase the value of PROMs, recall bias is a factor that ought to be accounted for when assessing patient reported outcomes [4,14].

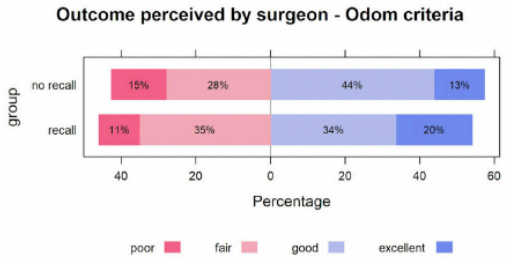

Fig. 2. Outcome perceived by the treating surgeon for each subgroup as measured by the criteria of Odom.

Lingard et al. reported statistically significant worse recalled preoperative pain scores for patients submitted to total hip arthroplasty when comparing to prospectively collected scores before surgery. Gotlin and colleagues demonstrated that retrospectively collected PROMs scores for patients submitted to arthroscopic rotator cuff surgery were significantly worse than the actual results collected preoperatively [19,20].

Pellisé et al. evaluated patient reported outcomes by comparing the results of 3 questionnaires fulfilled before posterior lumbar instrumentation surgery and postoperatively at a mean of 37.5 months after intervention and found that retrospectively assessing PROMs in those patients led to inaccurate PROMs recall and patient overestimation of surgical improvement [21]. Likewise, Dawson et al. had previously questioned the utility of using patient's recall of low back pain in retrospective studies as discrepancies regarding pain intensity were observed when patients were asked to evaluate low back pain severity 5 to 10 years after questionnaire fulfilment, although such discrepancies were not observed regarding pain location and functional limitation [22].

Even though the degree and extent of memory bias impact on PROMs may vary considerably those studies highlighted that relying on patients' recall of preoperative status can lead to inaccuracy when evaluating surgical improvement. Therefore, it can be hypothesized that postoperative evaluation of PROMs might as well be dented by patients' poor recall of preoperative status.

In this RCT, we aimed to establish whether recall bias also influenced postoperative scores by providing the patient intervention group their preoperative answers, and to our knowledge this is the first study to evaluate postoperative PROMs in this manner, as all previous studies have focused on comparing actual preoperative scores collected before surgery to patients' recall of health status many months or years after intervention.

Overall, higher EQ-5D and lower COMI and functional scores were observed in the intervention group, reflecting better clinical outcomes on PROMs in patients to whom the

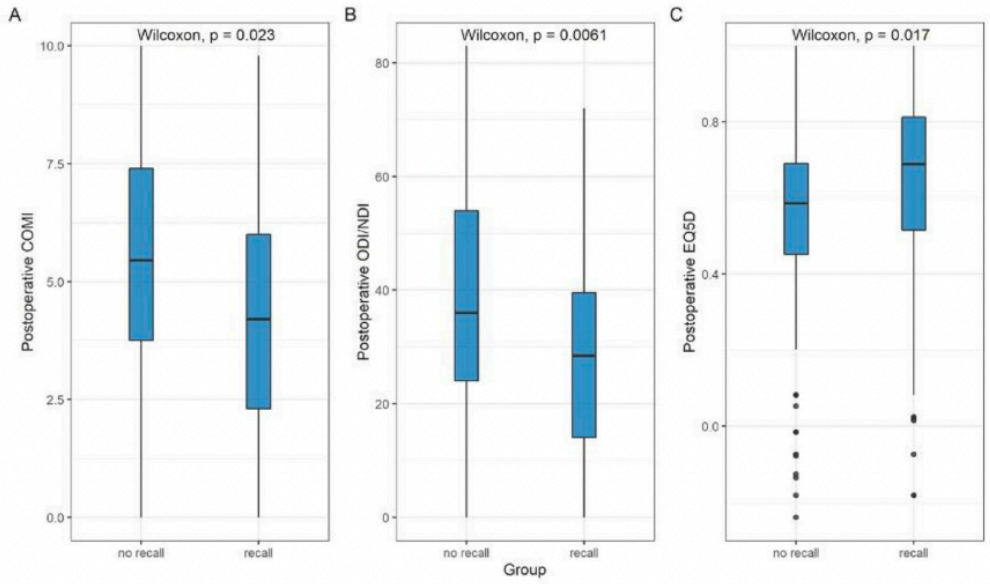

Fig. 3. Comparison of postoperative PROMs values between the groups, for overall patients. (Boxplots: the line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value, excluding outliers).

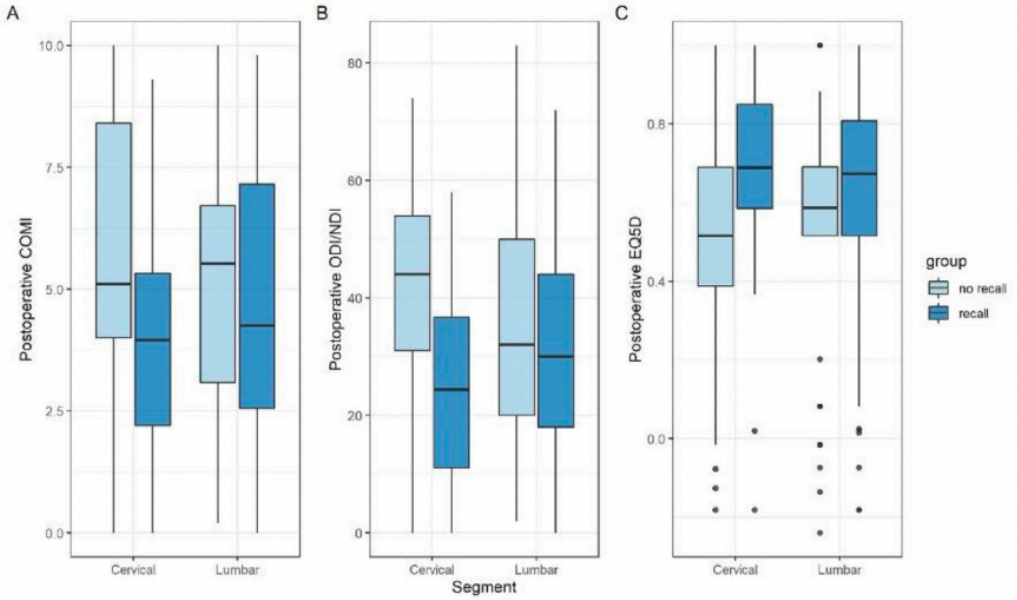

Fig. 4. Comparison of postoperative PROMs values between the groups and spine region, for overall patients.

preoperative answers were provided. These results were mainly reached due to the cervical patient subgroup. Of particular importance, clinically significant improvements occurred in the cervical intervention group for both COMI and NDI (Table 3) while in the cervical control group even the MCID for COMI was not accomplished.

In our population, both intervention and control groups had similar baseline clinical characteristics, namely regarding educational status, patient satisfaction with medical care and clinical outcomes, as measured per Odom's criteria. Hence, these results reinforce the belief that memory bias may impact postoperative scores, but contrary to many previous studies recall bias did not lead to overestimation of surgical results. Interestingly, in the study by Rodrigues et al. in a very similar population profile to ours, cervical spine patients (contrary to their lumbar counterparts) did not display a clinically significant memory bias of preoperative scores. This may signify that memory bias, at least within the median timeframe of both studies and for the PROMs employed both by us and Rodrigues et al. only alters scores to a certain extent and may explain why memory bias seemed to significantly impact only the results of lumbar patients in that study and of cervical patients in our study. Therefore, our results may translate a small, but in the case of cervical spine patients, clinically meaningful difference obtained by the correction of memory bias as it was not to be expected that recall bias within an average of 16.5 months after surgery would be so great as to lead to large discrepancies in PROMs scores in relation to the intervention group.

Several other reasons may explain our findings, namely patients' willingness to be sympathetic towards their physicians when confronted with pre-operative PROMs or patient perception that reporting less optimal scores may hinder the patient-doctor relationship.

Limitations

This study has some limitations; firstly, the lack of an objective measure of the outcome for comparison with the PROMs evaluated, therefore it cannot be guaranteed that the better postoperative scores obtained by the intervention group represent the most accurate results.

Secondly, it cannot be ruled out that patients in the intervention group may have been influenced by the provided preoperative scores and thus, anchoring bias might play a role in the differences between groups [23].

Additionally, providing patients their preoperative scores when fulfilling PROMs postoperatively can be burdensome in clinical practice which may limit its applicability on a daily basis.

This is also a single center study which limits its generalizability.

Nonetheless, our study tried to address a known issue associated with PROMs by focusing on the potential impact of memory bias in postoperative PROMs evaluation and

Table 2
Cervical patients' characteristics

	No recall (N=37)	Recall (N=29)	p value
Gender			.803*
Female	23 (62.2%)	19 (65.5%)	
Male	14 (37.8%)	10 (34.5%)	
Age			.892†
Median (Q1, Q3)	58.000 (49.000, 65.000)	58.000 (53.000, 64.000)	
Main clinical presentation			1.000*
cervical radiculopathy	13 (35.1%)	10 (34.5%)	
myelopathy	24 (64.9%)	19 (65.5%)	
BMI			.389†
Median (Q1, Q3)	27.300 (25.400, 29.940)	26.560 (24.753,28.175)	
Smoker			1.000*
yes	9 (24.3%)	6 (21.4%)	
no	28 (75.7%)	22 (78.6%)	
Education			.237*
≤4 years	14 (50.0%)	6 (30.0%)	
>4 years	14 (50.0%)	14 (70.0%)	
Preoperative COMI			.434†
Median (Q1, Q3)	8.1 (6.1, 8.8)	7.3 (6.2, 8.1)	
Preoperative NDI			.160†
Median (Q1, Q3)	54 (32, 64)	44 (32, 55)	
Preoperative EQ5D			.075†
Median (Q1, Q3)	0.516 (-0.005, 0.587)	0.585 (0.516, 0.689)	
Postoperative COMI			.011†
Median (Q1, Q3)	5.1 (4, 8.4)	4 (2.2, 5.3)	
Postoperative ODI/NDI			< .001†
Median (Q1, Q3)	44 (31, 54)	24.4 (11, 36.7)	
Postoperative EQ5D			.015†
Median (Q1, Q3)	0.516 (0.388, 0.690)	0.689 (0.585, 0.850)	
Satisfaction (COMI)			.430†
Median (Q1, Q3)	5 (5, 5)	5 (5, 5)	
Outcome (COMI)			.034†
Median (Q1, Q3)	3.5 (3, 4)	4 (4, 5)	
Odom			.791†
poor	5 (13.5%)	2 (6.9%)	
fair	11 (29.7%)	12 (41.4%)	
good	19 (51.4%)	11 (37.9%)	
excellent	2 (5.4%)	4 (13.8%)	
COMI improvement			.015†
Median (Q1, Q3)	1.2 (-0.5, 4.4)	3.4 (1.4, 4.4)	
NDI improvement			.026†
Median (Q1, Q3)	5 (-4.1, 20)	17 (6.8, 29)	
EQ5D improvement			.655†
Median (Q1, Q3)	0.069 (-0.052, 0.439)	0.104 (0.000, 0.275)	

* Fisher Exact Test for Count Data.
† Kruskal-Wallis rank sum test.

larger prospective studies are warranted to confirm whether giving patients their preoperative scores really can lead to a more accurate interpretation of postoperative results.

Interpretation

Lower COMI, NDI/ODI and higher EQ-5D scores were obtained for patients who were shown their preoperative answers, with clinically and statistically meaningful differences reached for cervical spine patients. Therefore, mitigating the impact of memory bias by providing patients with preoperative results upon postoperative questionnaires fulfilment influenced PROMs results in our study.

Prospective studies with larger series are needed to confirm these findings as the medical community strives to understand and establish the most adequate way to employ PROMs in clinical practice.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

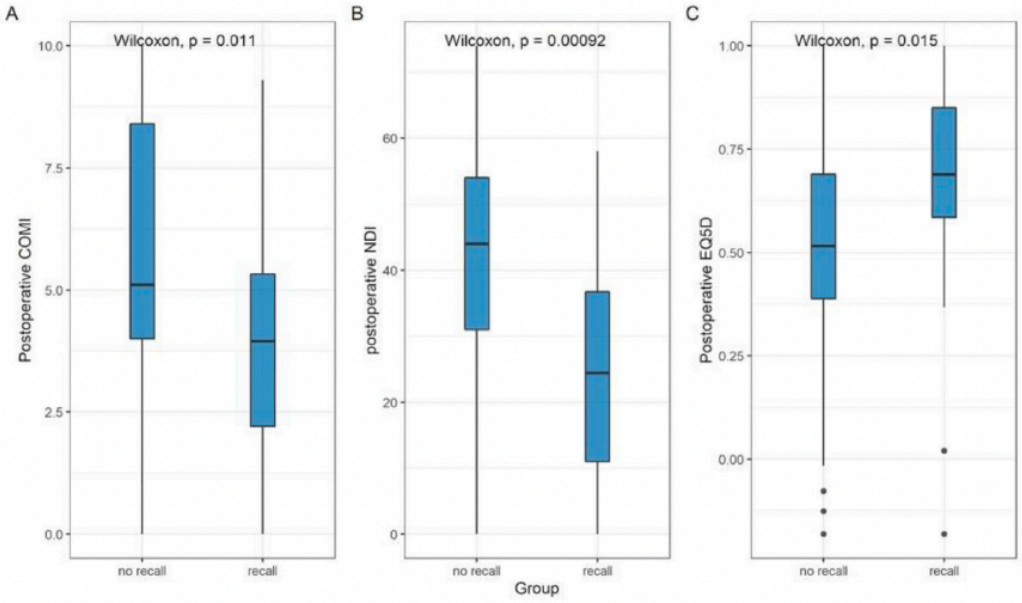

Fig. 5. Comparison of postoperative PROMs values between the groups, for patients with cervical spine pathology.

Table 3
PROMs improvement among groups

	Cervical no recall (N=37)	Cervical with recall (N=29)	Lumbar no recall (N=37)	Lumbar with recall (N=44)	MCID
COMI improvement Median (Q1, Q3)	1.2 (-0.5,4.4)	3.4 (1.4, 4.4)	2.7 (1, 4.5)	3 (0.8, 5.1)	2.6
ODI/NDI improvement Median (Q1, Q3)	5 (-4.2, 20)	17 (6, 29)	18 (10, 36)	15.6 (5, 31)	ODI: 14.9 NDI: 17.3
EQ5D improvement Median (Q1, Q3)	0.069 (-0.052, 0.439)	0.104 (0.000, 0.275)	0.175 (-0.033, 0.532)	0.242 (0.000, 0.559)	Lumbar: 0.49 Cervical: 0.24

Acknowledgment

This study was not supported by any kind of funding.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.spinee.2022.11.007.

References

[1] McCormick JD, Werner BC, Shimer AL. Patient-reported outcome measures in spine surgery. J Am AcadOrthop Surg 2013;21:99–107. https://doi.org/10.5435/JAAOS-21-02-99.

[2] Nayak NR, Coats JM, Abdullah KG, Stein SC, Malhotra NR. Tracking patient-reported outcomes in spinal disorders. Surg Neurol Int 2015;6 (Suppl 19):S490–9. https://doi.org/10.4103/2152-7806.166892.

[3] Basch E, Deal AM, Kris MG, Scher H I, Clifford AH, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016;34:557–65. https://doi.org/10.1200/JCO.2015.63.0830.

[4] Baeksted C, Pappot H, Nissen A, Hjollund NH, Mitchell SA, Basch E, et al. Feasibility and acceptability of electronic symptom surveillance with clinician feedback using the patient-reported outcomes version of common terminology criteria for adverse events (PRO-CTCAE) in Danish prostate cancer patients. J Patient Rep Outcomes 2017;1:1. https://doi.org/10.1186/s41687-017-0005-6.

[5] Girgis A, Durcinoska I, Levesque JV, Gerges M, Sandell T, Arnold A, et al. ehealth system for collecting and utilizing patient reported outcome measures for personalized treatment and care (PROMPT-Care) among cancer patients: mixed methods approach to evaluate feasibility and acceptability. J Med Internet Res 2017;19:e330. https://doi.org/10.2196/jmir.8360.

[6] Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-

Reported Outcome Measures (PROMs) in routine cancer care. J Med Radiat Sci 2021;68:186–95. https://doi.org/10.1002/jmrs.421.

[7] Aleem IS, Currier BL, Yaszemski MJ, Poppendeck H, Huddleston P, Eck J, et al. Do cervical spine surgery patients recall their preoperative status? Clin Spine Surg 2018;31:E481–7. https://doi.org/10.1097/bsd.0000000000000726.

[8] Aleem IS, Duncan J, Ahmed AM, Zarrabian M, Eck J, Rhee J, et al. Do lumbar decompression and fusion patients recall their preoperative status? Spine 2017;42:128–34. https://doi.org/10.1097/brs.0000000000001682.

[9] Rodrigues R, Silva PS, Cunha M, Vaz R, Pereira P. Can we assess the success of surgery for degenerative spinal diseases using patients' recall of their preoperative status? World Neurosurg 2018;115:e768–73. https://doi.org/10.1016/j.wneu.2018.04.174.

[10] Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Int Med 2010;152:726–32.

[11] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309:814–22. https://doi.org/10.1001/jama.2013.879.

[12] Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, et al. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18(Suppl 3):374–9. https://doi.org/10.1007/s00586-009-0931-y.

[13] Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine 2011;14:598–604. https://doi.org/10.3171/2010.12.SPINE10472.

[14] 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: clinical article. J Neurosurg Spine 2013;18:154–60. https://doi.org/10.3171/2012.10.SPINE12312.

[15] Pereira M, Cruz EB, Domingues L, Duarte S, Carnide F, Fernandes R. Responsiveness and interpretability of the portuguese version of the neck disability index in patients with chronic neck pain undergoing physiotherapy. Spine (Phila Pa 1976) 2015;40:E1180–6. https://doi.org/10.1097/BRS.0000000000001034.

[16] Broekema AEH, Molenberg R, Kuijlen JMA, Groen RJM, Reneman MF, Soer R. The odom criteria: validated at last. J Bone Joint Surg 2019;101:1301–8. https://doi.org/10.2106/jbjs.18.00370.

[17] Damasceno LHF, Rocha PAG, Barbosa ES, Barros CAM, Canto FT, Delfino HLA, 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;21:1273–82. https://doi.org/10.1007/s00586-011-2100-3.

[18] Mannion AF, Impellizzeri FM, Leunig-Jeszenszy D, Becker HJ, Haschtmann D, Preiss S, et al. FULL PAPER AWARD: time to remove our rose-tinted spectacles: a candid appraisal of the relative success of surgery in over 4500 patients with degenerative disorders of the lumbar spine, hip or knee. Eur Spine J 2018;27 778–8. https://doi.org/10.1007/s00586-018-5469-4.

[19] Gotlin MJ, Kingery MT, Baron SL, McCafferty J, Jazrawi LM, Meislin RJ. Recall bias in retrospective assessment of preoperative patient-reported American shoulder and elbow surgeons scores in arthroscopic rotator cuff repair surgery. Am J Sports Med 2020;48:1471–5. https://doi.org/10.1177/0363546520913491.

[20] Lingard EA, Wright EA, Sledge CB. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. J Bone Joint Surg Am 2001;83:1149–56. https://doi.org/10.2106/00004623-200108000-00003.

[21] Pellisé F, Vidal X, Hernández A, Cedraschi C, Bagó J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. Spine 2005;30:365–8. https://doi.org/10.1097/01.brs.0000152096.48237.7c.

[22] Dawson EG, Kanim LEA, Sra P, Dorey F, Goldstein TB, Delamarter RB, et al. Low back pain recollection versus concurrent accounts. Spine 2002;27:984–93. https://doi.org/10.1097/00007632-200205010-00020.

[23] Yasseri T, Reher J. Fooled by facts: quantifying anchoring bias through a large-scale experiment. J Comput Soc Sc 2022;5:1001–21. https://doi.org/10.1007/s42001-021-00158-0.

3.2 PAPER B

European Spine Journal
https://doi.org/10.1007/s00586-023-08093-6

ORIGINAL ARTICLE

Cross-cultural adaptation, validation and establishment of the minimal clinically important change score of the European Portuguese core outcome measures index in patients with lumbar degenerative spine disease

Pedro Valente Aguiar[1,2] · Pedro Santos Silva[1,2] · Diana Lucas[1,2] · Rui Vaz[1,2,3] · Paulo Pereira[1,2,3] · Anne F. Mannion[4]

Received: 24 May 2023 / Revised: 15 October 2023 / Accepted: 8 December 2023

Abstract

Purpose The core outcome measures index (COMI) for the back is a questionnaire that evaluates five domains and has been translated into several languages and validated for different populations. We aimed to translate, cross-culturally adapt and validate it in European Portuguese for use in patients with degenerative lumbar disease. Additionally, we aimed to establish the minimal clinically important change score (MCIC).

Methods The translation and cultural adaptation were done according to published guidelines. Patients awaiting surgery at a neurosurgical center completed the COMI, Oswestry Disability Index (ODI), EQ-5D questionnaires and a pain visual analog scale (VAS). To evaluate COMI's reproducibility, patients completed the questionnaire twice within two weeks, preoperatively, in addition to answering a transition question. The MCIC was determined by analysis of postoperative changes in total COMI score, using the anchor method, with a question ascertaining surgical outcome as perceived by the patient.

Results The first set of questionnaires was answered by 108 patients and the second, by 98 patients. COMI's construct validity was confirmed by demonstrating the hypothesized correlation between each domain's score (Spearman Rho > 0.4) and the corresponding questionnaire score (ODI, EQ-5D and VAS) and through adequate correlation (Spearman > 0.6) between COMI's total score and ODI and EQ-5D total scores. Intraclass correlation coefficients between each domain and COMI's total score were > 0.8. The MCIC was calculated as 2.1.

Conclusion The cross-culturally adapted COMI questionnaire is a valid clinical assessment tool for European Portuguese-speaking patients with degenerative lumbar disease, with an MCIC of 2.1 points.

Keywords COMI for the back · Degenerative spine disease · PROMs · European Portuguese · MCIC

Introduction

Patient-reported outcome measures (PROMs) are fundamental to evaluate the outcome of spine surgery and require the use of solid and reliable assessment tools [1, 2]. The core outcome measures index (COMI) is a simple and effective questionnaire that assesses five domains (pain, function, well-being, quality of life and disability) and is the PROM of choice in the Spine Tango surgical and conservative registry created by EUROSPINE, the Spine Society of Europe [3, 4].

Damasceno et al. [5] cross-culturally adapted and validated the COMI (back) for use in Brazilian–Portuguese-speaking patients within the Brazilian population; however, not only does Brazilian–Portuguese display relevant differences to European Portuguese but also both South American and European populations differ markedly regarding ethnical diversity and cultural habits [6, 7].

Therefore, the aim of this study was to cross-culturally adapt the COMI (back) to European Portuguese and to assess its reliability and validity in the Portuguese

✉ Pedro Valente Aguiar
pedrovalenteaguiar@protonmail.com

1 Neurosurgery Department, Centro Hospitalar Universitário São João, Alameda Prof Hernâni Monteiro, 4200-319 Porto, Portugal

2 Faculty of Medicine, Porto University, Porto, Portugal

3 Neurosciences Centre, Hospital da CUF, Porto, Portugal

4 Department of Teaching, Research and Development, Spine Division, Schulthess Klinik, Zurich, Switzerland

Published online: 05 January 2024

population. Additionally, we aimed to define a minimal clinically important change score (MCIC) for this PROM that would henceforth help standardize the interpretation of meaningful clinical improvement following surgery or conservative treatment in patients with degenerative diseases of the lumbar spine.

Materials and methods

Study design and population

The study was approved by the ethics committee of our institution in November 2021, and all patients gave written consent before taking part in the study.

From January to December 2022, consecutive adult patients awaiting surgery for degenerative diseases of the lumbar spine at a neurosurgical unit of a tertiary care hospital were asked to complete the preoperative COMI. Patients with non-degenerative lumbar disease such as infections, tumors or fractures were excluded. Patients answered the questionnaire twice before surgery within a 5- to 15-day time period: firstly in the outpatient clinic or by email and secondly upon arrival at the department before surgery. The time frame between both COMI assessments was chosen based on the premise defined by Terwee et al. [8] that the time between observations should be not too long such that clinical status may have altered and not too short to rule out recall bias.

An additional question enquiring about any change in status (including "no change," "slight change" and "great change" as possible answers) was asked before the second completion of the PROM to evaluate any alteration of the clinical status during that time period. "No change" was considered as displaying no significant change in clinical status.

Together with the COMI, patients also answered the Portuguese version of the Oswestry Disability Index (ODI) questionnaire to evaluate disability associated with lumbar degenerative disease [9]; the Portuguese 5 response category version of the European quality of life 5-dimension questionnaire (EQ-5D), to assess quality of life [10]; and a visual analogue scale (VAS), to analyze both back and leg pain from a 0 to 10 scale. The ODI scores can range from 0 to 100, with higher scores meaning worse clinical status, while for the EQ-5D, higher scores represent better quality of life (range − 0.536 to 1 for the Portuguese version) [11].

Core outcome measures index (back)

The COMI is made up of seven questions and has a maximum total score of 10 and a minimum of 0, with higher scores indicating worse clinical status. Scores are obtained from the average of the five domain scores: pain (taking the higher of the back pain and leg pain score, scored 0–10), function, well-being, quality of life and disability (taking the average of social and work disability scores). Function, well-being, quality of life and disability are rated on a 1–5 scale and must be converted into a 0–10 format by the formula (score − 1) × 2.5 before averaging with the pain score [12].

Creation of the new Portuguese version of the COMI (back)

We followed the guidelines established by Beaton et al. [13] to create the European Portuguese version of the COMI.

The initial English to Portuguese translation was done by the first author of this article, who is a native Portuguese speaker and has a degree in English language and philology, and a professional native Portuguese-speaking English translator, and a final consensus was reached for a Portuguese version. This version was translated back to the original language by two native English-speaking, bilingual (Portuguese) teachers and translators, both blind to the original COMI version and working independently.

Finally, and after the back translation was reviewed by an English-speaking spine research specialist with vast experience validating the COMI questionnaire, a consensus was reached to create the pre-final version by the four individuals that had performed the translations, another bilingual linguist and a neurosurgeon.

The pre-final version was tested in 27 patients with degenerative conditions of the lumbar spine, and as none reported any issues regarding their comprehension of the questionnaire, no alterations were made to this version, and it became final.

The same translation committee also performed the translation of four additional questions that were included to create the postoperative questionnaire at our department. In similar fashion to the English version available at the EUROSPINE website (https://www.eurospine.org/quality-assurance/spine-tango-registry/ressources/clinical-forms/), these four questions evaluate whether there were postoperative complications, repeat lumbar surgery and assess patient satisfaction with both medical care and surgical outcome (Table 1).

The supplementary file displays the new preoperative and postoperative versions used at our department that include the translated COMI.

Statistical analysis

Validity and reproducibility

Floor effects (worst status, highest COMI scores) and ceiling effects (best status, lowest COMI scores) indicate the

Table 1 Comparison between test–retest assessments of the COMI domains, overall score, ODI, EQ-5D and VAS, preoperatively

COMI first assessment		COMI second assessment	
		Change in clinical status	
		No change	93 (95%)
		Slight change	5 (5%)
		N-missing	10
COMI pain		*COMI pain*	
Mean (SD)	8.4 (1.5)	Mean (SD)	8.2 (1.6)
		N-missing	10
COMI back-related function		*COMI back-related function*	
Mean (SD)	7.4 (2.0)	Mean (SD)	7.6 (1.8)
		N-missing	10
COMI symptom-specific well-being		*COMI symptom-specific well-being*	
Mean (SD)	8.7 (1.9)	Mean (SD)	8.7 (1.8)
		N-missing	10
COMI quality of life		*COMI quality of life*	
Mean (SD)	7.4 (1.9)	Mean (SD)	7.3 (2.0)
		N-missing	10
COMI disability	6.1 (3.3)	*COMI disability*	6.5 (3.4)
Mean (SD)		Mean (SD)	
		N-missing	10
COMI overall score		*COMI overall*	
Mean (SD)	7.6 (1.5)	Mean (SD)	7.6 (1.5)
		N-missing	10
ODI overall score		*ODI overall score*	
Mean (SD)	58.8 (17.0)	Mean (SD)	57.9 (16.5)
		N-missing	12
EQ-5D overall score		*EQ-5D overall score*	
Mean (SD)	0.2 (0.2)	Mean (SD)	0.2 (0.2)
		N-missing	12
VAS pain (maximum value from either back % leg)		*VAS (maximum value from either back % leg)*	
Mean (SD)	8.1 (1.5)	Mean (SD)	8.1 (1.6)
N-miss	1	*N*-miss	12

number of patients at both extremes of clinical status for whom relevant deterioration or improvement, respectively, cannot be obtained and were calculated for each item of the COMI, being considered adverse if > 70% and ideal if < 15% [10].

Assessment of the COMI construct validity was done by formulating hypotheses based on the results of previous studies validating the same questionnaire in their respective language. Spearman rank correlation coefficients were used to evaluate the relationship between the COMI total score and each domain score and scores for the respective outcome domains of other validated questionnaires. Spearman ρ coefficients of 0–0.20 were considered poor, 0.21–0.40 fair, 0.41–0.60 good, 0.61–0.80 very good and 0.81–1 excellent [7, 14–16]. The COMI pain domain was hypothesized to correlate with the VAS for pain, the COMI function and disability domains with the ODI and the COMI well-being and quality of life domains with the EQ-5D, all with a $\rho > 0.4$. Moreover, the total COMI score was hypothesized to correlate with the total ODI and EQ-5D scores with a $\rho > 0.6$ [14].

Reproducibility gauges whether patients give consistent answers upon repeated completions of the questionnaire, and regarding COMI's reproducibility, it was hypothesized that the intraclass correlation coefficients (ICC) would be > 0.8 for each COMI domain and for the COMI total score. ICCs between 0.6 to 0.8 and > 0.8 reflect good and excellent reproducibility, respectively [10, 14].

Minimum detectable change (MDC) and minimal clinically important change score (MCIC)

The minimal detectable change score (MDC) equates to the minimum real change in a patient's PROM score that does not arise from measurement error. Using a confidence level

of 95%, we calculated the MDC based on the following formula: 2.77 × standard error of measurement (SEM) [10].

The MCIC represents a change in patient PROM score that translates to a relevant clinical improvement as perceived by physician or patient. No consensus exists regarding the most precise method to calculate its value, with both distribution-based and anchor-based methods having been proposed [5, 16]. We used an anchor-based method to calculate the MCIC with the "Global Treatment Outcome" item from the Spine Tango Patient Self-assessment Form as a reference to assess patient change in clinical status following surgery. The question "How much did the operation in our hospital help your back problem?" had possible five answers: "helped a lot," "helped," "helped only little," "didn't help" and "made things worse." To calculate the MCIC, we used "helped" as the minimum answer for depicting clinical improvement (Table 2).

With the use of receiver operating characteristic curve (ROC) analysis, the MCIC of the PROM was calculated as the threshold change in PROM score that most accurately distinguished patients reporting clinical improvement. This method was used to identify a threshold for the change in PROM score that maximized Youden's Index (which functionally equates to the sum of the sensitivity and specificity) or minimized the sum of false positive and false negative rates. R software (R Foundation for Statistical Computing, Vienna, Austria) version 4.0.3 was used for data analysis.

Table 2 Patient scores in the postoperative questionnaire

COMI overall score—post-op	
Mean (SD)	5.1 (2.4)
N-missing	29
ODI overall score—post-op	
Mean (SD)	30.3 (17.9)
N-missing	32
EQ-5D overall score—post-op	
Mean (SD)	0.5 (0.3)
N-missing	30
Improvement	
It did not help	7 (9.0%)
It helped only little	10 (12.8%)
It helped	22 (28.2%)
It helped a lot	39 (50.0%)
It made things worse	0 (0%)
N-missing	30

Results

Patient characteristics

During the time frame of the study, 170 patients underwent surgery due to degenerative disease of the lumbar spine in our department; those with questionnaires that were not completed properly were excluded.

A total of 108 patients (43 males and 65 females; mean age 56.4 years (SD: 11.9)) completed the newly adapted COMI on the first assessment. The mean preoperative COMI score was 7.6 (SD: 1.5), and the main patient complaint was sciatica (69.4%) followed by neurogenic claudication (28.7%).

Floor and ceiling effects

No adverse floor or ceiling effects were observed for any of the COMI domain scores. All ceiling effects were below 15%, while no floor effects were above 70%, with the highest floor effect being reached for the symptom-specific well-being domain at 59.2% (Table 3).

Construct validity

All pre-defined hypotheses regarding the construct validity of the European Portuguese version of the COMI were confirmed (Table 4), with the COMI pain domain scores having an excellent correlation with those of the VAS for pain ($\rho = 0.85$), while the scores for the domains function ($\rho = 0.49$), quality of life ($\rho = 0.54$) and disability ($\rho = 0.58$) each showed a good correlation with their corresponding scores on the ODI and EQ-5D. The well-being domain ($\rho = 0.401$) had a fair correlation with the EQ-5D. The total COMI score had very good correlation with the Oswestry

Table 3 Floor and ceiling effects

Domain	Ceiling effect (best status, lowest score) *N* (% of 108)	Floor effect (worst status, highest score) *N* (% of 108)
COMI pain	3 (0.93%)	30 (27.8%)
COMI back-related function	1 (0.93%)	24 (22.2%)
COMI symptom-specific well-being	1 (0.93%)	64 (59.26%)
COMI quality of life	6 (5.56%)	25 (23.15%)
COMI disability	9 (8.33%)	31 (28.70%)

Table 4 Correlations (Spearman's ρ) between COMI domain scores, the respective VAS, ODI and EQ-5D values and the total scores

Item 1	Item 2	Spearman's ρ
COMI pain	VAS	0.85
COMI function	ODI	0.50
COMI symptom-specific well-being	EQ-5D	− 0.40
COMI quality of life	EQ-5D	− 0.54
COMI disability	ODI	0.58
COMI total score	ODI total score	0.63
COMI total score	EQ-5D total score	− 0.70

Table 5 Test–retest reliability as measured by intraclass correlation coefficients

Domain	ICC (95% confidence interval)
COMI pain	0.78 (0.70–0.85)
COMI function	0.69 (0.58–0.78)
COMI symptom-specific well-being	0.84 (0.77–0.88)
COMI quality of life	0.81 (0.74–0.87)
COMI disability	0.86 (0.80–0.90)
COMI total score	0.89 (0.84–0.92)

Disability Index total score ($\rho = 0.63$) and the EQ-5D score ($\rho = -0.70$).

Reproducibility and reliability

To evaluate test–retest reliability, 98 patients completed the COMI questionnaire twice within a 15-day time period. Ninety-three patients reported no change between assessments while five reported a slight change in clinical status and were excluded from the ICC analysis. There was very good consistency in the answers obtained between the two assessment time points as defined by the intraclass correlation coefficients (Table 5). Excellent ICCs were observed for the domains well-being (ICC: 0.84), quality of life (ICC: 0.81) and disability (ICC: 0.86), whereas good ICCs were observed for pain (ICC: 0.78) and function (0.69). The total COMI score displayed an excellent ICC for its repeated completion (ICC: 0.89).

Minimum detectable change (MDC)

The standard of error measurement (SEM) for the total COMI score was calculated as 0.549; therefore, the MDC calculated according to the 2.77 × SEM formula was 1.52.

Minimal clinically important change (MCIC) score

Seventy-nine patients (73%) answered the postoperative questionnaire, within a mean postoperative time of 12 weeks.

The MCIC score determined from the ROC analysis was 2.1 with a sensitivity of 69% and a specificity of 94% (Fig. 1).

Discussion

Patient-reported outcome measures are very important tools at the disposal of clinicians and patients alike that help assess and document treatment outcome from the patient's perspective [16–18]. Therefore, their validity and reproducibility are of utmost importance.

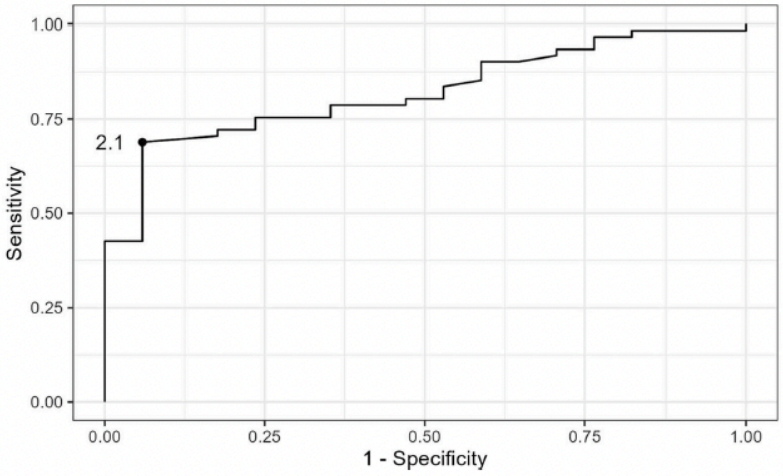

Fig. 1 MCIC obtained from the ROC curve (cutoff of 2.1 points associated with a sensitivity of 69% and specificity of 94%)

Due to its simplicity and reliability, the core outcome measures index is a useful PROM for evaluating the surgical outcome of patients undergoing spine surgery, allowing for standardization in the measurement of clinical outcomes [14, 16].

The COMI has been cross-culturally adapted and validated for use in numerous languages, but clinicians in Portugal have hitherto been using a Brazilian–Portuguese version, which is not culturally adapted to the European Portuguese-speaking population [7, 14]. Unsurprisingly, the new European Portuguese COMI version includes various expressions and words that differ substantially from the Brazilian version [supplementary material]. Additionally, to analyze clinically important improvement after surgery, spine surgeons have been using an MCIC that was defined neither for a Portuguese questionnaire nor for the Portuguese population, namely that based on the work by Mannion et al. on a German-speaking population undergoing spine surgery in Switzerland [16].

The psychometric properties of the European Portuguese version of the COMI were confirmed as being comparable with those reported in the literature for other languages, including the Brazilian–Portuguese version [5]. Regarding floor and ceiling effects, and despite the fact that our patients were all awaiting surgery and would therefore tend to display a more extreme clinical status than patients being managed conservatively, we found all domains to be within the ideal range, with symptom-specific well-being closer to having an adverse value, following a similar trend to its performance in other language versions, namely the Brazilian–Portuguese, Korean, Spanish and Polish versions [3, 5, 14, 20].

The construct validity for the European Portuguese COMI was also in line with that of most previous language versions of the COMI, with most domains and total scores having good to excellent correlation with their corresponding full-length questionnaires. Symptom-specific well-being displayed the lowest Spearman ρ value with the corresponding EQ-5D, which was also the case for the French ($\rho = -0.36$) and Korean ($\rho = -0.56$) versions. The test–retest reliability of the European Portuguese COMI was confirmed, with excellent ICCs being observed for all domains but two, namely function and pain, which nonetheless demonstrated good test–retest reliability.

The MDC for the COMI score in our work was 1.52, which is higher than the value for the Korean version (0.97) but slightly lower than that reported for the Hungarian (1.63), Brazilian–Portuguese (1.7), Polish (1.79) and German (1.74) versions. The MCIC of 2.1 was very similar to the value of 2.2 obtained by Mannion et al. [16] in another European spine center, highlighting the consistent psychometric properties of the COMI in different languages and populations [14, 21].

Limitations

One limitation of our study is that it was conducted in a single center, which might limit its generalizability. However, the Portuguese population is fairly homogenous and relatively small, and therefore, meaningful differences are not to be expected between spine centers [22].

Additionally, all patients included in our work were awaiting surgery, which might have led to a higher symptom burden than in a general outpatient population; however, this still did not lead to detrimental floor and ceiling effects and allowed for a homogenous group of patients to define the MCIC.

Conclusion

In this study, we evaluated the psychometric properties of a cross-culturally adapted European Portuguese version of the COMI in patients with lumbar degenerative spine disease. The new language version demonstrated strong reliability and construct validity, and its MDC and MCIC were established. The questionnaire can henceforth be used by spine surgeons for their European Portuguese-speaking patients to assess surgical outcome.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00586-023-08093-6.

Acknowledgements We would like to thank the translators involved in the cross-cultural adaptation and the office clerk at our institution for the important contribution to this work.

Declarations

Conflict of interest There are no conflicts of interest for this manuscript.

References

1. McCormick JD, Werner BC, Shimer AL (2013) Patient-reported outcome measures in spine surgery. J Am Acad Orthop Surg 21:99–107
2. Nayak NR, Coats JM, Abdullah KG, Stein SC, Malhotra NR (2015) Tracking patient-reported outcomes in spinal disorders. Surg Neurol Int 6(19):S490–S499. https://doi.org/10.4103/2152-7806.166892
3. Ferrer M, Pellise F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2006) Validation of a minimum outcome core set in the evaluation of patients with back pain. Spine Phila Pa (1976) 31(12):1372–1379. https://doi.org/10.1097/01.brs.0000218477.53318.bc
4. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective. Part 1: the core outcome

measures index in clinical practice. Eur Spine J 18(3):367–373. https://doi.org/10.1007/s00586-009-0942-8
5. Damasceno LH, Rocha PA, Barbosa ES, Barros CA, Canto FT, Defino HL, Mannion AF (2012) 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 21(7):1273–1282. https://doi.org/10.1007/s00586-011-2100-3
6. Silva AS (2008) Integrando a variação social e métodos quantitativos na investigação sobre linguagem e cognição: para uma sociolinguística cognitiva do português europeu e brasileiro. Rev Est Ling 16(1):49–81. https://doi.org/10.17851/2237-2083.16.1.49-81
7. Feldman-Bianco B (2001) Brazilians in Portugal, Portuguese in Brazil: constructions of sameness and difference. Identities Glob Stud Cult Power 8(4):607–650. https://doi.org/10.1080/1070289X.2001.9962710
8. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
9. Martins NS (2002) Adaptação Cultural e Linguística do Owestry Low Back Pain Disability Index – ODI 2.0 [Monografia]. Coimbra: Escola Superior de Tecnologia da Saúde de Coimbra
10. Ferreira PL, Ferreira LN, Pereira LN (2013) Contributos para a Validação da Versão Portuguesa do EQ-5D [Contribution for the validation of the Portuguese version of EQ-5D]. Acta Med Port 26(6):664–675
11. Ferreira PL, Antunes P, Ferreira LN, Pereira LN, Ramos-Goñi JM (2019) A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. Quality life Res Int J quality life Asp Treat care Rehabilit 28(12):3163–3175. https://doi.org/10.1007/s11136-019-02226-5
12. Pochon L, Kleinstück FS, Porchet F, Mannion AF (2016) Influence of gender on patient-oriented outcomes in spine surgery. Eur Spine J 25(1):235–246. https://doi.org/10.1007/s00586-015-4062-3
13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine Phila Pa (1976) 25(24):3186–3191
14. Kim HJ, Yeom JS, Nam Y, Lee NK, Heo YW, Lee SY, Park J, Chang BS, Lee CK, Chun HJ, Mannion AF (2018) Validation and cross-cultural adaptation of the Korean version of the core outcome measures index in patients with degenerative lumbar disease. Eur Spine J 27(11):2804–2813. https://doi.org/10.1007/s00586-018-5759-x
15. Matsudaira K, Oka H, Oshima Y, Chikuda H, Taniguchi Y, Matsubayashi Y, Kawaguchi M, Sato E, Murano H, Laurent T, Tanaka S, Mannion AF (2018) Development of the Japanese core outcome measures index (comi): cross-cultural adaptation and psychometric validation. BMC Musculoskelet Disord 19(1):71. https://doi.org/10.1186/s12891-018-1986-x
16. Mannion AF, Porchet F, Kleinstück FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: part 2 minimal clinically important difference for improvement and deterioration as measured with the core outcome measures index. Eur Spine J 18(3):374–379. https://doi.org/10.1007/s00586-009-0931-y
17. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, Rogak L, Bennett AV, Dueck AC, Atkinson TM, Chou JF, Dulko D, Sit L, Barz A, Novotny P, Fruscione M, Sloan JA, Schrag D (2016) Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 34(6):557–565. https://doi.org/10.1200/JCO.2015.63.0830
18. Nguyen H, Butow P, Dhillon H, Sundaresan P (2021) A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. J Med Radiat Sci 68(2):186–195. https://doi.org/10.1002/jmrs.421
19. Abdeldaiem A, Saweeres ESB, Shehab-Eldien AA, Mannion AF, Rehan Youssef A (2020) Cross-cultural adaptation and validation of the Arabic version of the core outcome measures index for the back (COMI-back) in patients with non-specific low back pain. Eur Spine J 29(10):2413–2430. https://doi.org/10.1007/s00586-020-06530-4
20. Miekisiak G, Kollataj M, Dobrogowski J, Kloc W, Libionka W, Banach M, Latka D, Sobolewski T, Sulewski A, Nowakowski A, Kiwic G, Pala A, Potaczek T (2013) Cross-cultural adaptation and validation of the Polish version of the core outcome measures index for low back pain. Eur Spine J 22(5):995–1001. https://doi.org/10.1007/s00586-012-2607-2
21. Klemencsics I, Lazary A, Valasek T, Szoverfi Z, Bozsodi A, Eltes P, Fekete TF, Varga PP (2016) Cross-cultural adaptation and validation of the Hungarian version of the core outcome measures index for the back (COMI Back). Eur Spine J 25(1):257–264. https://doi.org/10.1007/s00586-014-3750-8
22. OECD/European Observatory on Health Systems and Policies (2021) Portugal: Country Health Profile 2021, State of Health in the EU, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels https://health.ec.europa.eu/system/files/2021-12/2021_chp_pt_english.pdf.

3.3 PAPER C

European Spine Journal
https://doi.org/10.1007/s00586-024-08564-4

ORIGINAL ARTICLE

Cross-cultural adaptation, validation, and establishment of the minimal clinicallyimportant change score of the European Portuguese Core Outcome Measures Index in patients with cervical degenerative spine disease

Pedro Valente Aguiar[1,2] · Paulo Pereira[1,2] · Anne F. Mannion[3] · Pedro Santos Silva[1,2]

Received: 15 August 2024 / Revised: 22 October 2024 / Accepted: 10 November 2024

Abstract

Purpose We aimed to validate and cross-culturally adapt the Core Outcome Measures Index (COMI) neck for use in Portuguese patients with cervical spine degenerative disease and define the minimal clinically important change score (MCIC) for this questionnaire and population.

Methods The COMI translation and cross-cultural adaptation was done following published guidelines. Patients awaiting surgery in a neurosurgical centre completed the COMI, Neck Disability Index (NDI) and EQ-5D questionnaires, a pain visual analog scale (VAS) twice within a 5-to-15-day period, and a Global Treatment Outcome (GTO) question evaluating whether a clinical status change had occurred during that period. The MCIC was obtained through an anchor method by analysis of changes in pre- to postoperative total COMI scores and GTO dichotomized answers.

Results The COMI first assessment was completed by 101 patients and 72 patients completed both assessments. The questionnaire showed good construct validity ($n=72$ patients) as predefined hypotheses were confirmed: scores on each COMI domain correlated with a Spearman $\rho > 0.4$ with scores for the corresponding domain on other questionnaires and COMI total score displayed good correlation with total NDI score ($\rho=0.67$) and EQ-5D total score ($\rho=0.55$). Test-retest reliability ($n=72$ patients) was confirmed through high intraclass correlation coefficients. The MCIC ($n=76$ patients) was calculated as 2 points.

Conclusion The COMI (neck) psychometric qualities were confirmed, such that it can be considered a valid and reliable questionnaire to be applied in the European Portuguese population with surgical cervical spine degenerative disease, with an MCIC of 2 points.

Keywords COMI neck · Cervical spine degenerative disease · PROMs · European Portuguese · MCIC

Introduction

Cervical spine degenerative disease has a high prevalence and incidence in the population, with 24.2% of healthy individuals reported to have cervical spinal cord compression on imaging studies, of which an estimated 10% may develop myelopathy [1]. Given the high frequency of this condition and population ageing, it is expected that surgery will remain a powerful treatment option for patients who develop symptoms of myelopathy and / or radiculopathy [1, 2].

Assessing surgical outcome through reliable and effective patient reported outcome measures (PROMs) remains a fundamental part of clinical practice [3, 4]. The Core

✉ Pedro Valente Aguiar
pedrovalenteaguiar@protonmail.com

1 Neurosurgery Department, ULS São João– University Hospital, Alameda Prof Hernâni Monteiro, Porto 4200-319, Portugal

2 Faculty of Medicine, Porto University, Porto, Portugal

3 Department Teaching, Research and Development, Spine Division, Schulthess Klinik, Zurich, Switzerland

Published online: 22 November 2024

Outcome Measures Index (COMI) is a simple questionnaire that evaluates five domains (pain, function, symptom-specific well-being, quality of life, and disability) and is the reference PROM in the Spine Tango surgical and conservative registry created by EUROSPINE, the Spine Society of Europe [4]. This questionnaire has recently been cross-culturally adapted and validated for use in the European Portuguese population with lumbar degenerative spine disease [5]. However, the version of the COMI used in patients with cervical spine degenerative disease has never been translated and validated for use in the European Portuguese population.

The aim of this study was to cross-culturally adapt the COMI neck to European Portuguese and evaluate its reliability and validity in the Portuguese population with degenerative cervical spine disease. Additionally, we intended to define the minimal clinically important change score (MCIC) for this PROM, in order to contribute to the standardized measurement of meaningful clinical improvement obtained by this subgroup of patients and help physicians and patients better understand and define successful surgical outcome.

Materials and methods

Study design and population

The ethics committee of our institution approved this study in November 2021 and all patients were informed about the study and gave written consent before enrolment.

The study took place between January 2022 and April 2023 in a tertiary care Portuguese neurosurgical centre. Consecutive adult patients awaiting surgery for cervical spine degenerative disease completed the preoperative COMI on two different occasions. Firstly, patients answered the questionnaire by e-mail or phone and secondly, in the neurosurgical department upon arrival for surgery. The time interval between the two applications of the questionnaire varied between 5 and 15 days, as suggested by Terwee et al. [6], as fulfilment of the same questionnaire less than 5 days apart might lead to recall bias whereas after 15 days there is a higher chance of clinical status alteration. Indeed, before the second fulfilment of the questionnaire patients were asked about changes in clinical status since the first assessment (possible answers were: "no change", "slight change" and "great change") with only "no change" being considered as translating to no relevant alteration in clinical situation.

In addition to the COMI neck, patients were also asked to complete the European Portuguese version of the Neck Disability Index (NDI) [7, 8], the European Portuguese version of the European Quality of Life 5- dimension questionnaire (EQ-5D) [9], and a visual analogue scale (VAS), to report both neck and arm pain on a 0 to 10 scale. The NDI is a questionnaire that assesses neck related disability and is used to evaluate the condition of patients with degenerative cervical spine diseases, with scores ranging from 0 to 100 [7, 8] in which higher scores represent worse condition. The EQ-5D evaluates quality of life and scores can range from −0.536 to 1 for the European Portuguese version. Higher scores represent better quality of life [10].

The same patients who had completed the new preoperative COMI neck questionnaire were also asked to answer the new postoperative questionnaire during their outpatient visits, usually 6 to 15 weeks after surgery.

Patients were excluded from the study if they presented with non-degenerative pathologies such as tumours, infections, or fractures or if they showed cognitive impairment.

Core Outcome Measures Index (neck)

The COMI's score ranges from 0 to 10, with lower scores meaning better clinical status. Seven questions evaluating five domains make up the questionnaire with the COMI total score being the average of the scores from five domains: pain (using the higher of the neck or arm pain score, scored 0–10), function, well-being, quality of life, and disability (taking the average of social and work disability scores). Except for pain, the remaining domains are measured on a 1 to 5 format and converted into a 0 to 10 scale by the formula "(score − 1) × 2.5", before being averaged with the pain domain to give a final COMI score [11].

Creation of the new European Portuguese version of the COMI (neck)

The guidelines set by Beaton et al. [12] for translating this type of questionnaire were followed and the translation committee was made up of the same members who created the European Portuguese COMI (back) version [5]. The first translation was done by the first author of this manuscript (a physician with a university degree in English language teaching) and a native Portuguese with a professional background in English teaching and translation. A final consensus between the two translators was reached for the Portuguese version. This version was then translated back into English by two English native speakers with a professional background in Portuguese language teaching and translation. Both these translators worked independently and were blind to the original COMI version. Once the COMI (neck) version was created, it was revised by a spine research specialist with English language proficiency and many years of experience in validating PROMs. The final consensus for

the pre-final version of this PROM was obtained by the four previously mentioned translators, a bilingual linguist, and a bilingual neurosurgeon.

The pre-final version was then tested on 20 patients with degenerative cervical spine disease, and because no patients reported any issues with questionnaire comprehension, this version became final.

In addition to the COMI, the postoperative assessment of patients undergoing surgery due to cervical spinal degenerative pathology available at the EUROSPINE website (https://www.eurospine.org/quality-assurance/spine-tango-registry/ressources/clinical-forms/) includes 4 questions that evaluate postoperative complications, need for reintervention, patient satisfaction with medical care and Global Treatment Outcome. Those four questions were also translated into European Portuguese by the translation committee to create the final postoperative questionnaire.

[The supplementary file displays the new preoperative European Portuguese COMI (neck) and postoperative questionnaire that includes the COMI]

Statistical analysis

Validity and reproducibility

A ceiling effect occurs when patients record the best score on the scale and therefore no further improvement from any treatment is detectable (lowest COMI scores) and conversely a floor effect takes place when patients record the worst score on the scale such that deterioration after treatment cannot be detected (highest COMI scores). Both effects were calculated for each COMI item and were considered adverse if > 70% and ideal if < 15% [5, 12].

To assess the construct validity of the COMI (neck) we formulated hypotheses (see below) based on our previous validation of the COMI (back) and on multiple previous works validating this PROM to the respective languages. For that purpose, the relationship between the COMI total score and the respective domain score of other validated PROMs was evaluated using Spearman rank correlation coefficients. Spearman ρ coefficients of 0.81–1 were considered excellent, 0.61–0.80 very good, 0.41–0.60 good, 0.21–0.40 fair and 0–0.20 poor [5, 12–14].

The total COMI score was hypothesized to correlate with the total NDI and EQ-5D scores, each with a $\rho > 0.6$. The COMI pain domain was expected to correlate with the VAS for pain, the disability and function domains with the NDI, and the COMI well-being and quality of life domains with the EQ-5D; for each of these, the expected correlation was $\rho > 0.4$ [5, 12].

The test–retest reliability of a questionnaire reflects the ability of patients to provide consistent answers when completing a questionnaire on two different occasions. Excellent and good reproducibility is usually considered if the ICCs are > 0.8 and between 0.6 and 0.8, respectively [5, 12]. For the COMI (neck), it was hypothesized that the intraclass correlation coefficients (ICC) would be > 0.8 for the total score and for each domain of the COMI [6, 13].

Minimum detectable change (MDC) and minimal clinically important change score (MCIC)

The minimal detectable change score (MDC) for a PROM describes the smallest amount of true change in a patient´s condition that can be detected beyond that due to inherent measurement error and was calculated with a confidence level of 95% by using the fformula: 2.77 x standard error of measurement (SEM) [9].

The minimal clinically important change score (MCIC) reflects the change in the patient´s PROM score that represents a truly relevant clinical improvement as understood by patient and / or physician [5, 14].

Whilst there is no standardized way in the literature to obtain the MCIC, we utilized the same anchor-based method that had been used to obtain the MCIC for the COMI (back). This involves analysing the answers from the "Global Treatment Outcome" item from the Spine Tango Patient Self-Assessment Form to evaluate the change in patient clinical status after surgery. The item comprises the following question: "How much did the operation in our hospital help your neck problem?" and has 5 possible answers: "helped a lot", "helped", "helped only little", "didn´t help" and "made things worse". We considered meaningful improvement if patients answered "helped" or "helped a lot" to calculate the MCIC.

The MCIC was calculated using a receiver operating characteristic curve (ROC) analysis as.

the threshold change in PROM score that best distinguished patients reporting clinical improvement, maximising Youden's Index (sum of sensitivity and specificity) i.e. minimising the sum of false-positive and false-negative rates [5, 14, 15].

Data analysis was performed with version 4.0.3 of the R software (R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient characteristics

One-hundred and twenty-five patients underwent surgery due to degenerative cervical spine disease at our department from January 2022 to April 2023, of whom 24 had to be

Table 1 Mean values for the COMI items upon 1st assessment and floor and ceiling effects of the subdomain scores of the COMI

Domain	COMI 1st assessment ($n=101$) Mean (SD)	Ceiling effect (best status, lowest score) *N* (% of 101)	Floor effect (worst status, highest score) *N* (% of 101)
COMI pain	7.0 (2.6)	3 (2.9%)	17 (16.8%)
COMI neck-related function	6.2 (2.6)	4 (4%)	13 (12.9%)
COMI symptom specific well-being	8.7 (1.9)	1 (1%)	58 (57.4%)
COMI quality of life	7.5 (1.7)	25 (24.8%)	23 (22.8%)
COMI disability	6.1 (3.7)	10 (9.9%)	34 (33.7%)

Table 2 Correlations (Spearman´s ρ) between COMI domain scores, the respective visual analogue scale (VAS), neck disability index (NDI) and EQ-5D values and the total scores

Item 1	Item 2	Spearman´s ρ
COMI pain*	Pain VAS*	0.95
COMI function	NDI	0.54
COMI symptom specific well being	EQ-5D	-0.40
COMI quality of life	EQ-5D	-0.44
COMI disability	NDI	0.47
COMI total score	NDI total score	0.67
COMI total score	EQ-5D total score	-0.55

* the higher of neck and arm/shoulder pain score

excluded due to invalid questionnaire data. Questionnaires were considered invalid if not all items were answered in the COMI, or if more than one answer per item was given for any of the three evaluated questionnaires.

The first assessment of the COMI was completed by 101 patients (51 males and 50 females) with a mean (SD) age of 57.9 (13.4) years. The preoperative mean (SD) COMI score was 7.1 (1.7) and the most frequent pathology was myelopathy (73.3%) followed by radiculopathy (26.7%).

Floor and ceiling effects

Item analysis of the COMI demonstrated no relevant floor and ceiling effects (Table 1). The symptom specific well-being item of the COMI was the closest to having a floor effect, as 57.4% of patients reported being on the worst clinical status, but that value was well below the 70% threshold considered an undesirable effect.

Construct validity

The predefined hypotheses for each COMI domain were confirmed, as correlations with the respective comparator instrument of $\rho > 0.4$ were found for all domains: the pain domain had an excellent correlation with the pain VAS ($\rho = 0.95$); function and disability showed a good correlation with NDI ($\rho = 0.54$ and $\rho = 0.47$, respectively); quality of life showed good correlation with the EQ-5D ($\rho = -0.44$); and symptom specific well-being had a fair correlation with the EQ-5D ($\rho = -0.4$) (Table 2).

Construct validity was also confirmed for the COMI total score, displaying very good correlation with the total NDI score ($\rho = 0.67$) and good correlation with the EQ-5D total score ($\rho = -0.55$).

Reproducibility and reliability

Seventy-two patients completed both COMI assessments within the ideal 5-to-15-day time-period, and none reported an alteration in clinical status between assessments.

There was no significant difference between the mean COMI score on the first and the second assessment (7.1 (SD 1.7) and 7.2 (SD 1.7) respectively; $p = 0.88$). Very similar mean scores across the repeated assessments were also obtained for each individual COMI domain (Table 3).

The test-retest reliability of the questionnaire was confirmed, with excellent ICCs being obtained for the total COMI score (ICC: 0.87), pain (ICC: 0.9) and disability (ICC: 0.85) and good ICCs for function (ICC: 0.78),

Table 3 Test-retest reliability as measured by mean differences between the COMI total score and for each item and intraclass correlation coefficients

Domain	Mean (SD) score 1st assessment	Mean (SD) score 2nd assessment	*p*-value (matched pairs t-test)	ICC (95% confidence interval)
COMI pain ($n=72$)	7.0 (2.6)	7.1 (2.4)	0.94	0.90 (0.85–0.93)
COMI function ($n=72$)	6.3 (2.5)	6.6 (2.3)	0.53	0.78 (0.69–0.84)
COMI symptom specific well being ($n=72$)	8.5 (2.1)	8.5 (2.1)	0.84	0.79 (0.70–0.85)
COMI quality of life ($n=72$)	7.5 (1.6)	7.5 (1.9)	1	0.74 (0.64–0.82)
COMI disability ($n=72$)	6.4 (3.7)	6.5 (3.8)	0.94	0.85 (0.79–0.90)
COMI total score ($n=72$)	7.1 (1.7)	7.2 (1.7)	0.88	0.87 (0.82–0.91)
NDI total score ($n=67$)	47.6 (15.7)	48.4 (16.8)	0.7	
EQ-5D total score ($n=72$)	0.26 (0.24)	0.24 (0.25)	0.7	

Table 4 Patient scores in the postoperative questionnaire

COMI overall score_post-op	***N*=76**
Mean (SD)	5.1 (2.4)
N - missing	25
NDI overall score_post-op	***N*=32**
Mean (SD)	31.0 (12.5)
N - missing	69
EQ-5D overall score_post-op	***N*=36**
Mean (SD)	0.4 (0.3)
N - missing	65
Improvement	0 (0%)
It made things worse	9 (13.2%)
It did not help	0 (0%)
It helped only little	27 (39.7%)
It helped	32 (47.1%)
It helped a lot	33
N - missing	

symptom specific well-being (ICC: 0.79) and quality of life (ICC: 0.74) (Table 4).

Minimum detectable change (MDC)

The standard error of measurement was calculated as 0.637 which led to an MDC of 1.76.

Minimal clinically important change (MCIC) score

Seventy-six patients (75%) completed the postoperative COMI at a mean (SD) postoperative time of 12 (7.7) weeks (Table 4).

The ROC analysis was performed and a MCIC of 2 was obtained for a sensitivity of 66% and a specificity of 100% (Figure 1)

Discussion

Population ageing represents both a challenge and testament to the remarkable health strides achieved in recent decades [16]. Neifert et al. [17] and Kobayashi and colleagues [18] showed that the number of cervical spine surgeries has drastically increased in recent decades in two ageing countries, respectively the United States of America and Japan, with both authors predicting further significant increases in the number of spine surgeries in the upcoming decades. Portugal also has an ageing population [16] which will likely lead to increased numbers of surgeries due to degenerative cervical spine disease. Therefore, it is important to be able to assess surgical outcome using valid and reliable instruments that

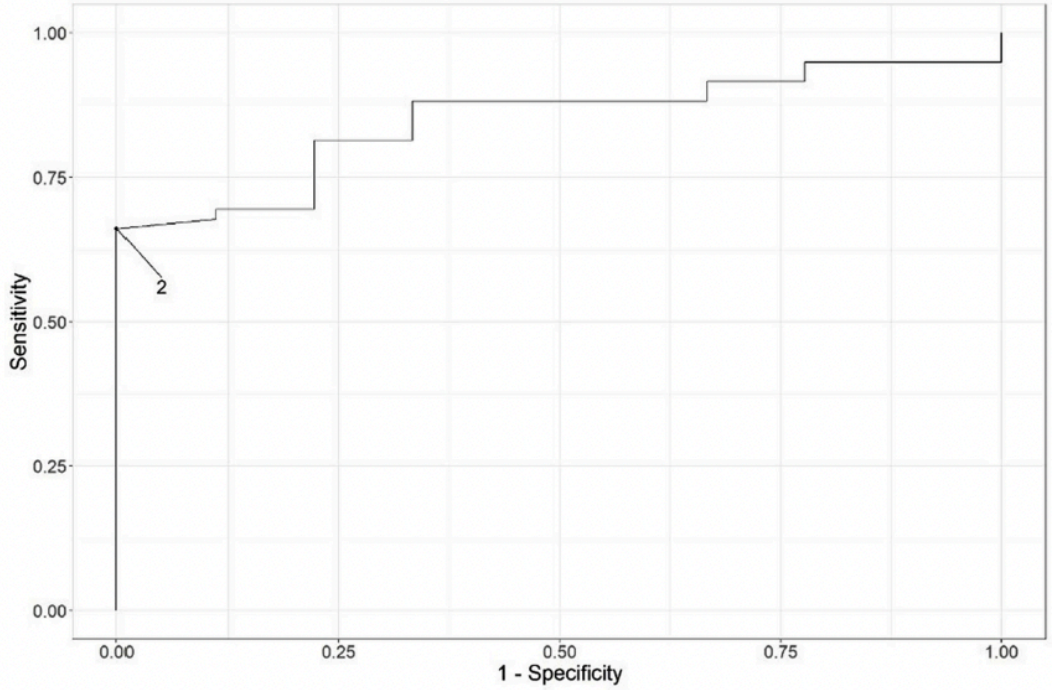

Fig. 1 MCIC obtained from the ROC curve (cut-off of 2 points associated with a sensitivity of 66% and specificity of 100%)

are adapted to the study population. Regarding spine pathology, the COMI (back or neck) is one of the most important and frequently used tools to evaluate treatment outcome in patients with lumbar and cervical degenerative spine disease [4, 5, 19]. Though cervical and lumbar spine pathology can coexist [20], individuals presenting with cervical spine degenerative disease represent a distinct subgroup of patients, with degenerative cervical myelopathy being the most frequent spinal cord disease and a potentially highly debilitating disease, with high societal burden [21].

With this study, we aimed to validate and cross–culturally adapt the COMI (neck) to the European Portuguese population with degenerative cervical spine disease and establish the minimal clinically important change for this subset of patients. Even though all our patients were awaiting surgery, and a higher symptom burden could be expected, we did not find any relevant floor or ceiling effects for each COMI item. The construct validity of the questionnaire was also confirmed, with good correlations with the comparator instruments for most items, except for the symptom specific well-being domain, which displayed only borderline correlation with the EQ-5D ($\rho = 0.4$). A similar trend was observed for the same domain in the validation of the European Portuguese COMI (back) ($\rho = 0.4$) [5], whereas some previous validations of the COMI (neck) displayed even poorer correlations between scores on this domain and the EQ-5D: Italian version ($\rho = -0.15$) and German version ($\rho = -0.268$) [19, 22]. Indeed, it is becoming increasingly clear that this item is less a measure of quality of life and more a measure of symptom severity, such that researchers may need to find an alternative comparator instrument with which to assess its validity. As for the total COMI score, very good correlation was found with the NDI total score ($\rho = 0.67$), while the correlation with the EQ-5D, albeit marginally below 0.6, was good ($\rho = -0.55$) and very similar to the German value ($\rho = -0.596$) [19]. The questionnaire also showed acceptable reliability, as measured by intraclass correlation coefficients that were either good (function, quality of life and symptom specific well-being) or excellent (total score, pain, disability).

We obtained an MDC of 1.76, which was comparable to previous COMI (neck) validations in Polish and Dutch, which reported MDC values of 1.97 and 1.7, respectively [23, 24], whereas during validation of the European Portuguese COMI (back) the MDC was 1.52 [5].

The MCIC of the European Portuguese COMI (neck) questionnaire was 2 points, similar to previously reported values: 2.1 for the European Portuguese COMI (back) [5] and Japanese COMI (neck) [25] and 2.2 for the German COMI (back) [14].

Limitations

This study validated the COMI neck in a subset of patients with myelopathy and / or radiculopathy awaiting surgery. As such, questions remain about the application of this questionnaire to a more general population with neck complaints or without surgical degenerative cervical spine pathology. However, the absence of ceiling and floor effects make us believe this version can broadly be applied to patients with degenerative cervical conditions.

Conclusion

For European Portuguese patients awaiting surgery for degenerative cervical spine disease, the COMI (neck) proved to be a valid and reliable questionnaire. MDC and MCIC values are presented that can help to define surgical success as perceived by patients and surgeons in this population.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00586-024-08564-4.

Acknowledgements We would like to thank the translators involved in the cross-cultural adaptation and the office clerk at our institution for the important contribution to this work.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Smith SS, Stewart ME, Davies BM, Kotter MRN (2021) The prevalence of asymptomatic and symptomatic spinal cord Compression on magnetic resonance imaging: a systematic review and Meta-analysis. Global Spine J 4597–607. https://doi.org/10.1177/2192568220934496
2. Gulati S, Vangen-Lønne V, Nygaard ØP, Gulati AM, Hammer TA, Johansen TO, Peul WC, Salvesen ØO, Solberg TK (2021) Surgery for degenerative cervical myelopathy: a Nationwide Registry-based observational study with patient-reported outcomes. Neurosurgery 15(4):704–711. https://doi.org/10.1093/neuros/nyab259
3. McCormick JD, Werner BC, Shimer AL (2013) Patient- reported outcome measures in spine surgery. J Am Acad Orthop Surg 21:99–107
4. Mannion AF, Porchet F, Kleinstuck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective. Part 1: the Core Outcome Measures Index in clinical practice. Eur Spine J 18(Suppl3):367–373. https://doi.org/10.1007/s00586-009-0942-8
5. Valente Aguiar P, Santos Silva P, Lucas D, Vaz R, Pereira P, Mannion AF (2024) Cross-cultural adaptation, validation and establishment of the minimal clinically important change score of the

European Portuguese core outcome measures index in patients with lumbar degenerative spine disease. Eur Spine J 2394–400. https://doi.org/10.1007/s00586-023-08093-6
6. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
7. Domingues L, Cruz E (2011) Relação entre a Catastrofização Da Dor, Perceção Da Intensidade Da Dor E Incapacidade Funcional em utentes com dor crónica cervical; Dissertação. de Mestrado em Fisioterapia; Relatório de Projeto de Investigação
8. Pereira M (2011) Contribuição para a adaptação cultural do Neck Disability Index E caraterização Da prática De Fisioterapia em pacientes com Dor Crónica Cervical. Universidade Nova de Lisboa
9. Ferreira PL, Ferreira LN, Pereira LN (2013) Contributos para a Validação Da Versão Portuguesa do EQ-5D [Contribution for the validation of the Portuguese version of EQ-5D]. Acta Med Port 26(6):664–675
10. Ferreira PL, Antunes P, Ferreira LN, Pereira LN, Ramos-Goñi JM (2019) A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. Quality of life research. Int J Qual life Aspects Treat care Rehabilitation 28(12):3163–3175. https://doi.org/10.1007/s11136-019-02226-5
11. Pochon L, Kleinstück FS, Porchet F, Mannion AF (2016) Influence of gender on patient-oriented outcomes in spine surgery. Eur Spine J 25(1):235–246. https://doi.org/10.1007/s00586-015-4062-3
12. Kim HJ, Yeom JS, Nam Y, Lee NK, Heo YW, Lee SY, Park J, Chang BS, Lee CK, Chun HJ, Mannion AF (2018) Validation and cross-cultural adaptation of the Korean version of the Core Outcome measures Index in patients with degenerative lumbar disease. Eur Spine J 27(11):2804–2813. https://doi.org/10.1007/s00586-018-5759-x
13. Matsudaira K, Oka H, Oshima Y, Chikuda H, Taniguchi Y, Matsubayashi Y, Kawaguchi M, Sato E, Murano H, Laurent T, Tanaka S, Mannion AF (2018) Development of the Japanese Core Outcome measures Index (COMI): cross-cultural adaptation and psychometric validation. BMC Musculoskelet Disord 19(1):71. https://doi.org/10.1186/s12891-018-1986-x
14. Mannion AF, Porchet F, Kleinstück FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 18 Suppl 3(Suppl 3374–379. https://doi.org/10.1007/s00586-009-0931-y
15. Damasceno LH, Rocha PA, Barbosa ES, Barros CA, Canto FT, Defino HL, Mannion AF (2012) 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 21(7):1273–1282. https://doi.org/10.1007/s00586-011-2100-3
16. Rodrigues AM, Gregório MJ, Sousa RD, Dias SS, Santos MJ, Mendes JM, Coelho PS, Branco JC, Canhão H (2018) Challenges of Ageing in Portugal: data from the EpiDoC Cohort. Acta Med Port 31(2):80–93. https://doi.org/10.20344/amp.9817
17. Neifert SN, Martini ML, Yuk F, McNeill IT, Caridi JM, Steinberger J, Oermann EK (2020) Predicting trends in cervical spinal surgery in the United States from 2020 to 2040. World Neurosurg 141:e175–e181. https://doi.org/10.1016/j.wneu.2020.05.055
18. Kobayashi K, Sato K, Kato F, Kanemura T, Yoshihara H, Sakai Y, Shinjo R, Ohara T, Yagi H, Matsubara Y, Ando K, Nakashima H, Imagama S (2022) Trends in the numbers of spine surgeries and spine surgeons over the past 15 years. Nagoya J Med Sci 84(1):155–162. https://doi.org/10.18999/nagjms.84.1.155
19. Fankhauser CD, Mutter U, Aghayev E, Mannion AF (2012) Validity and responsiveness of the Core Outcome measures Index (COMI) for the neck eur. Spine J 21(3):561. https://doi.org/10.1007/s00586-011-1921-4
20. Morishita Y, Buser Z, D'Oro A, Shiba K, Wang JC (2018) Clinical relationship of degenerative changes between the cervical and lumbar spine. Asian Spine J 12(2):343–348. https://doi.org/10.4184/asj.2018.12.2.343
21. Benzel EC, Ghogawala Z (2013) Introduction: cervical spondylotic myelopathy. Neurosurg Focus 35(1). https://doi.org/10.3171/2013.5.FOCUS13211
22. Monticone M, Ferrante S, Maggioni S, Grenat G, Checchia GA, Testa M, Teli MG, Mannion AF (2014) Reliability, validity, and responsiveness of the cross-culturally adapted Italian version of the core outcome measures index (COMI) for the neck. Eur Spine J 23(4):863–872. https://doi.org/10.1007/s00586-013-3092-y
23. Gadjradj PS, Chin-See-Chong TC, Donk D, Depauw P, Tulder MWV, Harhangi BS (2021) Cross-cultural adaptation and psychometric validation of the Dutch Version of the Core Outcome measures Index for the Neck in patients undergoing surgery for degenerative disease of the cervical spine. Neurospine 18(4):798–805. https://doi.org/10.14245/ns.2142682.341
24. Miekisiak G, Banach M, Kiwic G, Kubaszewski L, Kaczmarczyk J, Sulewski A, Kloc W, Libionka W, Latka D, Kollataj M, Zaluski R (2014) Reliability and validity of the Polish version of the Core Outcome measures Index for the neck. Eur Spine J 23(4):898–903. https://doi.org/10.1007/s00586-013-3129-2
25. Oshima Y, Nagata K, Nakamoto H, Sakamoto R, Takeshita Y, Ohtomo N, Kawamura N, Iizuka M, Ono T, Nakajima K, Higashikawa A, Yoshimoto T, Fujii T, Tanaka S, Oka H, Matsudaira K (2021) Validity of the Japanese core outcome measures index (COMI)-neck for cervical spine surgery: a prospective cohort study. Eur Spine J 30:402–409. https://doi.org/10.1007/s00586-020-06657-4

4. DISCUSSION

This thesis provides new evidence that recall bias significantly alters postoperative PROM scores in patients undergoing cervical spine surgery — an underrecognized phenomenon with direct implications for clinical practice and data interpretation. This can significantly alter the way postoperative questionnaires will be applied in clinical practice. Moreover, we validated the COMI-back and COMI-neck to European Portuguese and cross-culturally adapted it to the Portuguese population with degenerative disease of the lumbar and cervical spine. This PROM showed good validity and consistency making it suitable for use in the European Portuguese population with degenerative spine disease. Additionally, surgical success from a patient´s perspective can now be assessed as a minimal clinically important change score was created for both the COMI-back and COMI-neck.

In the following chapter the main findings of this thesis are presented, namely through an article-by-article analysis and review. This discussion aims to go beyond a simple presentation and description of the results of the articles and aims to expand on those by framing our findings within a broader revision of the literature involving PROMs.

MAIN FINDINGS

4.1 Are the results of patient-reported outcome measures after spine surgery influenced by recall of preoperative scores? A randomized controlled trial

4.1.1 Background and aims of this work

Recall bias in patient-reported outcome measures has likely been overlooked or downplayed in clinical practice.

In 2018 and in the same tertiary Portuguese neurosurgical center where this PhD research took place, Rodrigues et al. assessed the significance of recall bias and its repercussion on patients' recollection of preoperative status.[62] They compared the actual preoperative PROM scores to a set of scores collected one year after surgery regarding patient recall of preoperative status. They found poor to moderate agreement between recalled and actual preoperative scores for all patient-reported outcome measures. However, these results were not similar across groups as patients who underwent surgery due to degenerative disease of the cervical spine showed good recall of preoperative status. Conversely, patients submitted to lumbar spine surgery had poor recall, particularly patients that reported good surgical outcome. [62]

4.1.2 Study design and methods

In a similar population, we conducted a randomized controlled trial to assess the impact of recall bias on the scores of postoperative PROMs in patients submitted to surgery due to degenerative disease of the spine.

To study this impact the patients of the intervention group received postoperative PROMs that included the preoperative answers for every question, whereas the control group answered postoperative PROMs without the preoperative responses. During preoperative and postoperative evaluation patients with lumbar degenerative spine disease completed the COMI-back and ODI questionnaires, whereas patients with cervical spine disease filled the COMI-neck and NDI forms. Both groups completed an overall health assessment through the EQ-5D. The primary endpoint was the median postoperative COMI score, and the secondary endpoints were the median ODI / NDI and EQ-5D scores.

4.1.3 Results

A total of 236 (96 cervical and 140 lumbar) patients were enrolled in this RCT. Valid postoperative PROM scores were obtained for 147 patients (74 from the control group and 73 from the intervention group), 88 (59.9%) females and 59 (40.1%) males with a median age of 58 years. The mean time between surgery and questionnaire fulfilment was 17 months and both groups shared similar baseline clinical characteristics.

The intervention group showed better outcomes than the control group, with lower median COMI, ODI, NDI scores, and higher EQ-5D scores. However, this improvement did not reach statistical significance in the lumbar patients.

Cervical patients

The better postoperative scores obtained by the intervention group were mainly due to the cervical subgroup of patients. Sixty-six patients with cervical pathology were included in the study, 37 (56.0%) from the control group and 29 (44.0%) from the intervention group and a detailed analysis of this surgical cohort was performed.

Primary endpoint (cervical patients)

- COMI: Median (IQR) was 3.95 (2.20−5.33) in the intervention group vs. 5.10 (4.00−8.40) in the control group (Wilcoxon, $p = 0.011$).

Importantly, only the intervention group achieved an improvement exceeding the MCID (2.6 as per Mannion et al. [71]) of the COMI, with an observed improvement of 3.4 compared to 1.2 in the control group.

Secondary endpoints(cervical patients)

- NDI: 24.40 (11.00−36.70) in the intervention group vs. 44.00 (31.00−54.00) in the control group (Wilcoxon, $p < 0.001$).
- EQ-5D: 0.689 (0.585, 0.850) in the intervention group vs. 0.516 (0.388, 0.690) in the control group (Wilcoxon, $p < 0.015$).

Additionally, item-level analysis revealed consistently better postoperative scores across all domains for patients who had access to their preoperative responses.

4.1.4 Discussion

To our knowledge, this is the first study to assess postoperative PROMs by providing patients with their original preoperative scores. Previous studies had focused on comparing actual preoperative scores collected before surgery to patients´ recall of health status many months or years after intervention.

Within the timeframe of our study, the findings indicate that recall bias may impact both preoperative perceived disability and postoperative perceived benefit if patients are not reminded of their preoperative state. Providing access to original preoperative scores significantly improved patients' assessments of postoperative outcome across multiple PROMs in patients with degenerative disease of the cervical spine. However, this was not observed in a statistically significant manner in the lumbar spine population. That contrasts with the study by Rodrigues et al. where memory bias of preoperative status was more pronounced in the population with lumbar degenerative disease. [62]

We hypothesized that recall error, at least within the median timespan of our study and for these questionnaires, only modestly alters PROM scores. This may explain why recall bias primarily affected lumbar patients in the study by Rodrigues et al. and cervical patients in our research.

Accordingly, our results may translate a small, but in the case of cervical spine patients, clinically meaningful difference obtained by the correction of recall bias. Indeed, it was not expected that recall bias within an average of 17 months after surgical intervention would be so major as to lead to large discrepancies in PROMs scores in relation to the control group.

We believe the clinically meaningful impact of recall bias on postoperative PROM scores is time-dependent from intervention and may be smaller or greater depending on this variable.

Other factors could have contributed to these results such as patients´ willingness to be sympathetic towards their physicians when confronted with preoperative PROMs or patient perception that reporting less optimal scores may hinder the patient-doctor relationship. Namely, being shown their own preoperative responses may have further heightened patients' awareness of the lack of anonymity in PROMs. This and because PROMs were collected by the treating surgeon may have reinforced socially desirable answering behavior by conditioning patients and reduced the spontaneity of PROM fulfilment.

Providing patients with their preoperative answers in postoperative PROMs can also increase administrative burden, which further highlights the need for new technologies to help streamline this integration.

Additionally, this was a single-center study which limits its generalizability to the non-Portuguese populations and other healthcare systems. Moreover, our follow-up duration (mean 17 months) may not capture the full trajectory of decline in memory accuracy over longer periods.

Nonetheless, there was a difference towards better outcomes in all patients in the intervention group and this difference reached clinical significance in the cervical cohort of patients. Therefore, this RCT demonstrated that recall bias impacts postoperative PROM scores, a factor that should not be overlooked.

In conclusion, and with a view to fully optimize the interpretation of PROMs in clinical practice and mitigate memory bias we recommend considering the integration of original preoperative scores when administering postoperative assessments. This process should ideally be supported by digital tools or electronic health record systems to minimize administrative burden and facilitate seamless implementation.

4.2 Cross-cultural adaptation, validation and establishment of the minimal clinically important change score of the European Portuguese Core Outcome Measures Index in patients with lumbar degenerative spine disease

4.2.1 Background and aims of this work

The accurate interpretation of patient-reported outcome measures in both research and clinical settings necessitates that these instruments undergo rigorous processes of cross-cultural adaptation and psychometric validation tailored to the target population. [9]

PROMs must be concise, comprehensible, and easily completed, ensuring minimal burden on patients while maintaining the integrity of the data collected. Moreover, patients should be made aware of the clinical value these tools represent, as their responses contribute to personalized and improved healthcare delivery. Consequently, the availability of thoroughly validated and culturally adapted instruments is essential. [9]

The Core Outcome Measures Index is a widely used, concise multidimensional PROM designed to evaluate key aspects of spinal disorders, including pain, function, symptom-specific well-being, quality of life, and disability. [29] Due to its brevity and comprehensiveness, the COMI has been adopted as the preferred PROM in several national spine registries, including the Spine Tango Registry managed by EUROSPINE, the European Spine Society. [29] Despite its widespread use and validation in several languages, prior to this study, a validated version of the COMI-back in European Portuguese was not available. Given the increasing implementation of PROMs in spine registries and clinical practice across Portugal and Europe, addressing this gap was both timely and necessary. Moreover, a minimal clinically important change score was defined for this questionnaire.

4.2.2 Study Design and Methods

Between January and December 2022, adult patients scheduled for surgery due to degenerative lumbar spine disease at a tertiary neurosurgical center completed the European Portuguese version of the COMI questionnaire. Patients with non-degenerative conditions were excluded. To assess test–retest reliability, patients completed the COMI twice before surgery within a 5- to 15-day interval, selecting only those reporting no clinical change for reliability analyses. Construct validity was evaluated by correlating COMI scores with other PROMs like the Oswestry Disability Index, EQ-5D, and visual analogue scales for back and leg pain.

The translation and cross-cultural adaptation followed established guidelines, involving independent forward and back translations, expert review, and pre-testing in patients to ensure clarity and cultural relevance.[8] Additional postoperative questions were also translated which led to the creation of the postoperative questionnaire applied to Portuguese patients with lumbar degenerative disease of the spine, that includes the COMI back.

4.2.3 Psychometric Evaluation

The psychometric properties of the European Portuguese COMI-back were assessed, focusing on floor and ceiling effects, construct validity, reproducibility, and responsiveness. Floor and ceiling effects were evaluated to ensure the instrument's sensitivity across clinical severity. Construct validity hypotheses were tested via Spearman correlations between COMI domains and corresponding validated instruments, anticipating moderate to strong correlations. Reproducibility was assessed using intraclass correlation coefficients (ICC), with values above 0.8 indicating excellent reliability. The minimal detectable change (MDC) was calculated to quantify the smallest change beyond measurement error, while the minimal clinically important change was established using an anchor-based approach with patient-reported global treatment outcomes. Receiver operating characteristic curve analysis determined the MCIC threshold that best distinguished clinical improvement.

4.2.4. Results

Between January and December 2022, 170 patients underwent surgery for degenerative lumbar spine disease. Incomplete questionnaires were excluded and 108 patients (43 males, 65 females; mean age 56.4 ± 11.9 years) completed the European Portuguese version of the COMI at baseline. The mean preoperative COMI score was 7.6 ± 1.5, with sciatica (69.4%) and neurogenic claudication (28.7%) as the primary complaints.

No significant floor or ceiling effects were observed, with all ceiling effects below 15% and the highest floor effect at 59.2% for symptom-specific well-being. Construct validity was confirmed, with excellent correlation between the COMI pain domain and VAS pain scores ($\rho = 0.85$), and good correlations for function ($\rho = 0.49$), quality of life ($\rho = 0.54$), and disability ($\rho = 0.58$) domains compared to ODI and EQ-5D. The total COMI score had strong correlation with the ODI ($\rho = 0.63$) and EQ-5D ($\rho = -0.70$) total scores.

Test–retest reliability was evaluated in 93 clinically stable patients, demonstrating excellent intraclass correlation coefficients for well-being (0.84), quality of life (0.81), disability (0.86), and total COMI score (0.89).

The minimum detectable change was 1.52, and the minimal clinically important change was established at 2.1 with 69% sensitivity and 94% specificity.

4.2.5 Discussion

This work presents the first cross-cultural adaptation and validation of the Core Outcome Measures Index for the European Portuguese-speaking population, filling a long-standing gap in spine outcomes research and clinical practice in Portugal. While the COMI had been extensively validated in multiple languages and settings, prior to this work, spine clinicians in Portugal were limited to using the Brazilian Portuguese version. [34] This limitation presented some challenges, not only for patient comprehension but also for accurate measurement and comparison of outcomes across populations. Although Brazilian Portuguese and European Portuguese share linguistic roots, differences in vocabulary, expressions, and cultural context can affect patient comprehension and response accuracy, making a dedicated European Portuguese version essential for valid assessment.

The psychometric validation presented in this study demonstrates that the European Portuguese version of the COMI maintains strong measurement properties, aligning with previous validations in other languages. [34,35,36] The observed correlations with established patient-reported outcome measures such as the ODI, VAS and EQ-5D support its construct validity. Moreover, the test–retest reliability was good for most domains, and the minimum detectable change and minimal clinically important change thresholds were well-defined, enabling its use both in clinical follow-up and research contexts.

Importantly, this study reinforces the central role of PROMs in patient-centered spine care. As tools designed to reflect the patient's subjective experience, PROMs must be linguistically and culturally adapted to the populations they assist. [33] Validating such tools ensures the consistency of outcome reporting, which in turn influences clinical decision-making, health policy, and research comparability. [9] The creation of a population-specific MCIC was particularly noteworthy, as it aligns the interpretation of PROM score changes to a level that is not only statistically but also clinically meaningful for patients.

We used an anchor-based method to define the MCIC because it relates score changes to an external criterion meaningful to patients. Unlike distribution-based methods, which are

statistically sound but may lack clinical relevance, the anchor-based approach ensured that clinical changes were meaningful from the patient's perspective.

Additionally, the absence of problematic floor and ceiling effects in this surgical cohort supports the strength of the COMI even in a population with severe symptom burden, such as patients awaiting surgery. The similarity of psychometric properties with other international validations, such as those from Poland, Brazil and Korea, emphasizes the reliability and versatility of this questionnaire in different populations with degenerative disease of the lumbar spine. [34,35,72]

In the context of this thesis, which aims to critically examine and improve the application of PROMs in degenerative spine surgery, this study presents a case example of how PROMs must be refined and contextualized for the populations they aim to serve. By offering a validated, culturally appropriate tool for the European Portuguese-speaking population, it strengthens the capacity for standardized outcome measurement, facilitates participation in international registries, and advances the spine field towards more precise and objective patient-reported outcome evaluation.

Moreover, focusing on a surgical cohort of patients allowed for a rigorous test of the instrument's sensitivity to change, culminating in the first empirically derived MCIC for a Portuguese-language COMI.

While the single-center nature of the study may limit generalizability, the relative linguistic and cultural homogeneity of the native Portuguese population partly mitigates this concern. [73] However, in the past few years Portugal saw a significant increase in its immigrant population. [74] This population is multicultural, and many immigrants may not have fluent control of European Portuguese, so this questionnaire will not be as valid and adapted to a significant portion of the immigrant population. [75] This highlights the need for further linguistic and cultural adaptations in Portugal.

Additionally, this study did not assess measurement invariance across different subgroups within the Portuguese population, such as by age, gender, or educational level through a multi-group confirmatory factor analysis. However, it must be noted that most existing validations of the COMI and similar PROMs have not incorporated measurement invariance testing. [34,35,72] As such, this remains a relevant gap in the psychometric evaluation of the COMI across different languages and versions.

In summary, this work represents a critical advancement in the application of PROMs in spine surgery in Portugal. It not only enables a more accurate and standardized assessment of degenerative lumbar spine surgical outcomes in European Portuguese speakers but also lays

the groundwork for future research aimed at refining PROM implementation and enhancing patient-centered care across diverse linguistic and cultural groups within Portugal.

4.3 Cross-cultural adaptation, validation, and establishment of the minimal clinically important change score of the European Portuguese Core Outcome Measures Index in patients with cervical degenerative spine disease

4.3.1 Background and aims of this work

Degenerative disease of the cervical spine is a common clinical entity with a growing impact on healthcare systems worldwide. [75] Imaging studies suggest that up to 24.2% of asymptomatic individuals exhibit cervical spinal cord compression, with close to 10% developing symptomatic myelopathy. [76] As populations continue to age, the number of patients requiring surgical intervention for cervical myelopathy and/or radiculopathy is expected to rise significantly. [77] In this context, accurately assessing treatment outcomes becomes essential for both clinical decision-making and quality control in spine care.

Patient-reported outcome measures are indispensable tools for capturing the patient's perspective on treatment outcomes. Among them, the Core Outcome Measures Index stands out as a fundamental questionnaire. [29]

As part of the scientific outcome of this thesis, a version tailored to lumbar spine pathology (COMI-back) had been successfully translated and validated for use in the European Portuguese population. However, no equivalent adaptation existed for degenerative cervical spine pathology — specifically, the COMI-neck.

Despite some similarities between cervical and lumbar degenerative conditions, the symptomatology, functional implications, and surgical approaches differ substantially. [77] Therefore, the existence of a validated COMI-back did not replace the need for a dedicated and culturally adapted COMI-neck version. The lack of such a questionnaire limits the ability of Portuguese-speaking physicians and researchers to evaluate surgical outcomes in patients with degenerative disease of the cervical spine in a standardized and internationally comparable manner.

The primary aim of this study was to cross-culturally adapt the COMI-neck for the European Portuguese population and to evaluate its psychometric properties, including reliability and validity. Furthermore, we defined the minimal clinically important change for this instrument, in order to establish a meaningful threshold for clinical improvement. Therefore, this work was created to support both clinical practice and research by validating a robust and culturally appropriate measure for outcome assessment in cervical spine surgery.

4.3.2 Study Design and Methods

The methodology of this study followed an identical pattern to the one used for creation of the COMI-back.

Consecutive adult patients scheduled for surgery due to cervical degenerative spine disease were recruited between January 2022 and April 2023. Patients completed the COMI-neck questionnaire preoperatively on two occasions, 5 to 15 days apart, to assess test–retest reliability. Only those reporting no change in clinical status between assessments were included in reliability analyses. In addition, participants completed the European Portuguese versions of the Neck Disability Index, EQ-5D, and visual analogue scales for neck and arm pain to allow for construct validation.

Postoperative outcomes were evaluated during outpatient follow-up (6 to 15 weeks post-surgery), when the same patients completed the COMI-neck again along with four standard postoperative questions from the Spine Tango registry. The questions cover complications, reintervention, patient satisfaction, and global treatment outcome to create the final postoperative version of the questionnaire that includes the new COMI-neck.

The cross-cultural adaptation process followed the guidelines of Beaton et al., involving forward translation, back-translation, expert committee review, and pretesting in 20 patients. [9] Since no comprehension issues were reported during pretesting, the pre-final version was adopted as the final version.

4.3.3 Psychometric Evaluation

Ceiling and floor effects were examined for each COMI item and considered acceptable if <15% and adverse if >70%. Construct validity was assessed through pre-established hypotheses and Spearman correlations between the COMI-neck and corresponding domains of the NDI, EQ-5D, and VAS. Correlations >0.6 were expected for total scores, and >0.4 for domain-specific comparisons.

Test–retest reliability was assessed using intraclass correlation coefficients, with values >0.8 considered excellent.

The minimal clinically important change was obtained via an anchor-based method using the “Global Treatment Outcome” item from the Spine Tango form. ROC curve analysis was used to define the score change best distinguishing improved patients, maximizing Youden’s Index.

4.3.4. Results

The questionnaires of 101 patients (mean age 57.9 years; 51 males, 50 females) were analyzed. The most common diagnosis was myelopathy (73.3%), followed by radiculopathy (26.7%). The mean preoperative COMI score was 7.1.

No relevant floor or ceiling effects were encountered. The symptom-specific well-being item approached a floor effect (57.4%) but remained below the 70% threshold.

The construct validity of the COMI was strong as all predefined hypotheses were confirmed. Strong correlations were found between the COMI domains and their respective comparator instruments:

Pain with VAS: $\rho = 0.95$ (excellent)

Function and disability with NDI: $\rho = 0.54$ and 0.47 (good)

Quality of life and well-being with EQ-5D: $\rho = -0.44$ and -0.40 (good to fair)

Total COMI with NDI: $\rho = 0.67$ (very good); with EQ-5D: $\rho = -0.55$ (good)

Regarding the reproducibility and reliability of this PROM, seventy-two patients completed both COMI assessments within 5–15 days, with no reported clinical changes. Mean scores remained stable (7.1 vs. 7.2; $p = 0.88$). ICCs confirmed excellent reliability for the total COMI (0.87), pain (0.90), and disability (0.85), and good reliability for function (0.78), well-being (0.79), and quality of life (0.74).

The minimum detectable change (MDC) was 1.76.

The minimal clinically important change, derived from the ROC analysis, was 2 points, with a sensitivity of 66% and specificity of 100%.

4.3.5. Discussion

Population ageing, while a testament to public health advances, presents growing challenges in the management of chronic and degenerative conditions, including those affecting the spine. [77]

Countries such as Japan and the United States have reported a marked increase in cervical spine surgeries in recent decades, a trend expected to continue as populations age. [79,80]

Portugal faces similar demographic shifts, which will likely translate into higher surgical volumes for degenerative cervical spine disease. [73] In this context, the availability of valid, reliable, and culturally adapted outcome instruments becomes essential to evaluate and guide clinical decision-making and healthcare resource allocation.

Cervical spine pathology, particularly degenerative cervical myelopathy, represents a unique clinical condition with significant potential for neurological impairment and societal burden. [75] Therefore, validating the COMI-neck for this subgroup is especially relevant.

Our study successfully achieved a cross-cultural adaptation and validation of the COMI- neck for the European Portuguese population undergoing surgery for degenerative cervical spine disease. Significantly, no relevant floor or ceiling effects were found, even though the population exhibited a high symptom burden preoperatively. This suggests the questionnaire's ability to capture a broad spectrum of clinical severity, supporting its utility across different stages of disease.

Construct validity was confirmed, with most domains correlating as expected with established PROMs. While the symptom-specific well-being domain demonstrated only a fair correlation with the EQ-5D ($\rho = 0.4$), this is consistent with earlier observations in other COMI validations. It may reflect a conceptual limitation of the EQ-5D as a comparator for this item, which appears to function more as a marker of symptom intensity rather than general quality of life. This finding suggests that future research may need to consider alternative instruments for validating this domain.

The total COMI score showed very good correlation with the NDI ($\rho = 0.67$) and good correlation with the EQ-5D ($\rho = -0.55$), aligning closely with results from other language versions and reaffirming the consistency of the COMI across cultural contexts. Reproducibility was also high, with intraclass correlation coefficients ranging from good to excellent across all domains.

With this study we set the minimal detectable change and minimal clinically important change for the COMI-neck. This is of particular importance because both MDC and MCIC are crucial for interpreting PROM data in both clinical and research settings. An MDC of 1.76 and an MCIC of 2.0 are congruent with those reported in other language versions, suggesting the consistency of these thresholds. Importantly, these values offer a practical benchmark for defining meaningful postoperative improvement in Portuguese patients undergoing cervical spine surgery. As with the validation of the COMI-back we opted for an anchor-method to define the MCIC because it aligns with patients´ perceived change in clinical status.

The limitations of this study resemble those of the lumbar version, including its single-center design, lack of assessment of measurement invariance across subgroups, and limited applicability to non-native or multicultural populations.

Nonetheless, this study thus provides the first validated and culturally appropriate version of the COMI-neck for European Portuguese speakers, offering clinicians and researchers in Portugal a reliable tool for assessing patient outcomes in degenerative cervical spine disease. As with the COMI-back, the widespread adoption of this PROM within spine care physicians in Portugal could enhance the standardization of outcome measurement, facilitate quality improvement initiatives, and contribute to more patient-centered care.

5. FINAL CONSIDERATIONS

This thesis presents an important contribution to the field of degenerative spine surgery by addressing a critical gap in the measurement and interpretation of patient-reported outcomes. In an era where patient-centered care and evidence-based medicine are increasingly prioritized, the accurate assessment of surgical success extends beyond traditional clinical or radiological parameters to incorporate the patient's own perspective on their pain, function, and quality of life.

This thesis focused on the improvement of methodological rigor when applying PROMs in clinical practice by incorporating and analyzing the impact of recall bias. Moreover, it provided a comprehensive validation of PROM instruments specific to spine care within the European Portuguese population, centering notably on the COMI questionnaire for both lumbar and cervical degenerative conditions.

However, the importance of this thesis to spine surgery care goes beyond the simple validation of questionnaires. As demographic trends worldwide point to an aging population with a rising incidence of degenerative spinal diseases, the demand for surgical interventions is expected to grow correspondingly. In Portugal, these shifts are anticipated to challenge healthcare providers and systems alike. The implementation of reliable, valid, and culturally adapted PROMs enables clinicians to better quantify patient outcomes, tailor treatments, and ultimately improve clinical decision-making. It also allows for the standardization of outcome measurement, which is fundamental to multicenter research, quality improvement programs, and health policy development.

This work's originality lies in several key areas. It offers a rigorous, culturally adapted version of the COMI questionnaire tailored to European Portuguese speakers, extending the applicability of this well-validated instrument into a previously underserved linguistic and cultural context. The process of cross-cultural adaptation, psychometric validation, and determination of minimal clinically important changes were performed following established international guidelines, ensuring that the final instruments possess both linguistic equivalence and clinical relevance.

Second, the thesis expands existing knowledge by directly and originally addressing recall bias in postoperative PROM assessments via a randomized controlled trial, providing valuable evidence on the temporal reliability of patient self-reporting after spine surgery. This is a crucial

methodological consideration often overlooked in spine outcome research but fundamental for the appropriate timing and interpretation of PROM data.

Third, the simultaneous validation of both the lumbar and cervical versions of the COMI in Portuguese offers a comprehensive toolkit for spine specialists, facilitating consistent evaluation across different degenerative spine pathologies. The demonstration of strong construct validity, excellent reproducibility, and defined thresholds for meaningful clinical improvement (MDC and MCIC) equips clinicians and researchers with practical tools to interpret patient outcomes confidently.

It is also noteworthy that in the validation studies of the COMI divergent validity was not directly addressed. It is a subtype of construct validity that assesses the degree to which a measurement instrument does not correlate strongly with measures of different, unrelated constructs. However, due to the COMI's brief, multidimensional design and overlap between domains, identifying truly unrelated constructs for divergent validity testing is challenging and uncommon. Thus, we prioritized confirming expected correlations with related PROMs through convergent validity, in similar fashion to multiple other validation studies.

Despite these strengths, this thesis has inherent limitations that warrant consideration. The sample sizes, while adequate for validation studies, remain relatively modest and drawn from a single tertiary care neurosurgical center. This may limit the generalizability of findings to broader or more diverse populations within Portugal or internationally.

We also conducted the randomized controlled trial on recall bias before the validation studies, which led to using an international MCID to gauge surgical success in both RCT groups. However, the MCID from Manion et al (2.6) did not differ substantially from the MCIC obtained for the COMI-back (2.1) and COMI-neck (2). Most importantly, it did not lead to clinically meaningful differences when comparing improvement between the intervention and control groups in the recall study.

Although our validation study of the European Portuguese version of the COMI-back and COMI-neck comprehensively addressed most of the COSMIN guidelines, we did not specifically evaluate structural validity and internal consistency. Structural validity is important for confirming the dimensional structure of the instrument and is typically analyzed through exploratory or confirmatory factor analysis. However, given the well-established factor structure of the COMI in various previous validations across diverse populations and considering that our study focused primarily on cross-cultural adaptation, reliability, and construct validity, we considered this to be beyond the scope of the current work. Similarly, internal consistency was not directly assessed in this thesis due to the COMI's brevity and

multidimensional nature (comprising distinct domains). Cronbach's alpha is a common measure of internal consistency, but it assumes one-dimensionality so is less informative and rigorous in the multidimensional context of the COMI.

Additionally, the focus on degenerative cervical and lumbar spine disease excludes other spine conditions such as deformity, trauma, or tumors, which may require different or additional PROM instruments. While the COMI is well suited for degenerative pathology, its applicability to other spine surgery indications remains to be established. [81]

The reliance on self-reported questionnaires also introduces inherent subjectivity, susceptibility to response bias, and potential limitations due to health literacy or cognitive impairments, particularly in elderly populations. Indeed, while the recall study and validation processes address some of these concerns by assessing memory bias, reliability and interpretability, many of the limitations of PROMs such as measurement invariance, health literacy, logistical challenges and subjectivity were not directly evaluated in this research.

Finally, the MCIC values determined in this thesis, while consistent with international reports, remain context-dependent and may vary with patient expectations, cultural factors, and healthcare delivery settings. Caution is advised when applying these thresholds universally without consideration of local nuances.

In conclusion, this thesis lays a strong foundation for the routine incorporation of PROMs into degenerative spine surgery care in Portugal. By validating culturally adapted and psychometrically sound instruments, it enhances the precision and relevance of patient outcome measurement and Portuguese patients can now receive clearer, more personalized feedback on their recovery.

The insights gained from this work also have the potential to inform clinical practice, support research projects, and ultimately improve patient-centered care and resource allocation in spine surgery. Future research should build upon these findings to expand PROM utilization across broader spine pathologies and diverse patient populations, integrating technological advances such as artificial intelligence and electronic PROM systems to streamline data collection and analysis.

The enduring value of this work is its alignment with the ongoing paradigm shift toward valuing the patient's perspective in evaluating treatment success—an essential step in delivering care that is not only effective but also meaningful to those who undergo it.

REFERENCES

[1] Berwick, D. M. (2015). Measuring surgical outcomes for improvement: Was Codman wrong? *JAMA, 313*(5), 469–470. https://doi.org/10.1001/jama.2015.4

[2] Churruca, K., Pomare, C., Ellis, L. A., Long, J. C., Henderson, S. B., Murphy, L. E. D., Leahy, C. J., & Braithwaite, J. (2021). Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expectations, 24*(4), 1015–1024. https://doi.org/10.1111/hex.13254

[3] Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., & Carr, A. J. (2010). The routine use of patient-reported outcome measures in healthcare settings. *BMJ, 340*, c186. https://doi.org/10.1136/bmj.c186

[4] Black, N. (2013). Patient reported outcome measures could help transform healthcare. *BMJ, 346*, f167. https://doi.org/10.1136/bmj.f167

[5] Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

[6] U.S. Department of Health and Human Services, Food and Drug Administration. (2009). *Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims*. https://www.fda.gov/media/77832/download

[7] European Medicines Agency. (2005). *Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products* (EMA/CHMP/EWP/139391/2004).

[8] Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

[9] Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

[10] Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: Bibliographic study of patient assessed health outcome measures. *BMJ, 324*(7351), 1417. https://doi.org/10.1136/bmj.324.7351.1417

[11] Deyo, R. A., & Weinstein, J. N. (2001). Low back pain. *New England Journal of Medicine, 344*(5), 363–370. https://doi.org/10.1056/NEJM200102013440508

[12] Fairbank, J. C., & Pynsent, P. B. (2000). The Oswestry Disability Index. *Spine, 25*(22), 2940–2952. https://doi.org/10.1097/00007632-200011150-00017

[13] Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, D., Brown, J. M., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology, 22*(4), 714–724. https://doi.org/10.1200/JCO.2004.06.078

[14] Barbera, L., Sutradhar, R., Seow, H., Earle, C. C., Howell, D., Mittmann, N., Li, Q., & Thiruchelvam, D. (2020). Impact of standardized Edmonton Symptom Assessment System use on emergency department visits and hospitalization: Results of a population-based retrospective matched cohort analysis. *JCO Oncology Practice, 16*(9), e958–e965. https://doi.org/10.1200/OP.20.00066

[15] Balitsky, A. K., Rayner, D., Britto, J., Lionel, A. C., Ginsberg, L., Cho, W., Wilfred, A. M., Sardar, H., Cantor, N., Mian, H., Levine, M. N., & Guyatt, G. H. (2024). Patient-reported outcome measures in cancer care: An updated systematic review and meta-analysis. *JAMA Network Open, 7*(8), e2424793. https://doi.org/10.1001/jamanetworkopen.2024.24793

[16] Bruce, B., & Fries, J. F. (2003). The Stanford Health Assessment Questionnaire: Dimensions and practical applications. *Health and Quality of Life Outcomes, 1*, 20. https://doi.org/10.1186/1477-7525-1-20

[17] Fries, J. F., Spitz, P., Kraines, R. G., & Holman, H. R. (1980). Measurement of patient outcome in arthritis. *Arthritis & Rheumatism, 23*(2), 137–145.

[18] Bruce, B., & Fries, J. F. (2005). The Health Assessment Questionnaire (HAQ). *Clinical and Experimental Rheumatology, 23*(5 Suppl 39), S14–S18. https://doi.org/10.1002/art.1780230202

[19] Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., Revicki, D., Symonds, T., Parada, A., & Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Quality of Life Research, 17*(2), 179–193. https://doi.org/10.1007/s11136-007-9295-0

[20] Falavigna, A., Dozza, D. C., Teles, A. R., Wong, C. C., Barbagallo, G., Brodke, D., Al-Mutair, A., Ghogawala, Z., & Rie, K. D. (2017). Current status of worldwide use of patient-reported outcome measures (PROMs) in spine care. *World Neurosurgery, 108*, 328–335. https://doi.org/10.1016/j.wneu.2017.09.002

[21] Mensah, E., Chalif, J. I., Baker, J. G., & Groff, M. W. (2024). Challenges in contemporary spine surgery: A comprehensive review of surgical, technological, and patient-specific issues. *Journal of Clinical Medicine, 13*(18), 5460. https://doi.org/10.3390/jcm13185460

[22] Roland, M., & Fairbank, J. (2000). The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine, 25*(24), 3115– 3124. https://doi.org/10.1097/00007632-200012150-00006

[23] Copay, A. G., Glassman, S. D., Subach, B. R., Berven, S., Schuler, T. C., & Carreon, L. Y. (2008). 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. *The Spine Journal, 8*(6), 968–974. https://doi.org/10.1016/j.spinee.2007.11.006

[24] Fairbank, J. C., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*(8), 271–273.

[25] van der Ploeg, R. J., Fidler, V., & Oosterhuis, H. J. (1996). Oswestry Disability Index: Translation into Spanish and validation in patients with low back pain. *Spine, 21*(24), 2825–2829. https://doi.org/10.1097/00007632-199612150-00020

[26] Mannion, A. F., Junge, A., Fairbank, J. C., Dvorak, J., & Grob, D. (2006). Development of a German version of the Oswestry Disability Index. Part 1: Cross-cultural adaptation, reliability, and validity. *European Spine Journal, 15(1)*, 55–65. https://doi.org/10.1007/s00586-004-0815-9

[27] Roland, M., & Morris, R. (1983). A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability. *Spine, 8(2)*, 141–144. https://doi.org/10.1097/00007632-198303000-00004

[28] Vernon, H., & Mior, S. (1991). The Neck Disability Index: A study of reliability and validity. *Journal of Manipulative and Physiological Therapeutics, 14(7)*, 409–415.

[29] Mannion, A. F., Junge, A., Fairbank, J. C., Dvorak, J., & Grob, D. (2006). Development of a core outcome measure for low back pain: COMI. *European Spine Journal, 15(Suppl 1)*, S54–S65. https://doi.org/10.1007/s00586-005-1044-8

[30] Chiarotto, A., Maxwell, L. J., Terwee, C. B., Wells, G. A., Tugwell, P., & Ostelo, R. W. (2016). Roland-Morris Disability Questionnaire and Oswestry Disability Index: Which Has Better Measurement Properties for Measuring Physical Functioning in Nonspecific Low Back Pain? Systematic Review and Meta-Analysis. *Physical therapy, 96(10)*, 1620–1637. https://doi.org/10.2522/ptj.20150420

[31] Davidson, M., & Keating, J. L. (2002). A comparison of five low back disability questionnaires: reliability and responsiveness. *Physical therapy, 82(1)*, 8–24. https://doi.org/10.1093/ptj/82.1.8

[32] Klasen, B., Hallmann, M., Schaub, S., Finger, T., & Mannion, A. F. (2007). Validation and reliability of the German version of the Core Outcome Measures Index (COMI) for the neck. *European Spine Journal, 16(12)*, 2105–2115. https://doi.org/10.1007/s00586-007-0508-4

[33] Mannion, A. F., Haschtmann, D., Puhakka, J., Kleinstück, F. S., Porchet, F., Loibl, M., Reitmeir, R., Jeszenszky, D., & Fekete, T. F. (2025). The Core Outcome Measures Index (COMI) comes of age: 18 years of efficient and effective patient-reported outcome measurement in the field of spinal disorders. Part 1. Development of the COMI, interpretation of scores, and use in research. *European Spine Journal*. Advance online publication. https://doi.org/10.1007/s00586-025-08835-8

[34] Damasceno LH, Rocha PA, Barbosa ES, Barros CA, Canto FT, Defino HL, Mannion AF (2012) 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, 21(7)*:1273–1282. https://doi.org/10.1007/s00586-011-2100-3

[35] Kim HJ, Yeom JS, Nam Y, Lee NK, Heo YW, Lee SY, Park J, Chang BS, Lee CK, Chun HJ, Mannion AF (2018) Validation and cross-cultural adaptation of the Korean version of the core outcome measures index in patients with degenerative lumbar disease. *Eur Spine J, 27(11)*:2804–2813. https://doi.org/10.1007/s00586-018-5759-x

[36] Matsudaira K, Oka H, Oshima Y, Chikuda H, Taniguchi Y, Matsubayashi Y, Kawaguchi M, Sato E, Murano H, Laurent T, Tanaka S, Mannion AF (2018) Development of the Japanese core outcome measures index (COMI): cross-cultural adaptation and psychometric validation. *BMC Musculoskelet Disord, 19(1)*:71. https://doi.org/10.1186/s12891-018-1986-x

[37] Cruz, E. B., Fernandes, R., Carnide, F., Domingues, L., Pereira, M., & Duarte, S. (2015). Cross-cultural adaptation and validation of the neck disability index to European Portuguese language. *Spine, 40(2)*, E77–E82. https://doi.org/10.1097/BRS.0000000000000692

[38] MacDermid, J. C., et al. (2009). Measurement properties of the Neck Disability Index: A systematic review. *J Orthop Sports Phys Ther, 39(5)*, 400–417. https://doi.org/10.2519/jospt.2009.2930

[39] Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol Group: Past, Present and Future. *Applied health economics and health policy, 15(2)*, 127–137. https://doi.org/10.1007/s40258-017-0310-5

[40] Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30(6)*, 473–483. https://doi.org/10.1097/00005650-199206000-00002

[41] Ferreira, P. L., Ferreira, L. N., & Pereira, L. N. (2013). Contributos para a validação da versão portuguesa do EQ-5D. *Acta Médica Portuguesa, 26(6)*, 664–675. https://doi.org/10.20344/amp.1317

[42] Ferreira, P. L. (2000). Development of the Portuguese version of MOS SF-36. Part II — Validation tests. *Acta Médica Portuguesa, 13(3)*, 119–127.

[43] Myers, C., & Wilks, D. (1999). Comparison of Euroqol EQ-5D and SF-36 in patients with chronic fatigue syndrome. *Quality of life research, 8(1-2)*, 9–16. https://doi.org/10.1023/a:1026459027453

[44] McDonough, C. M., Grove, M. R., Tosteson, T. D., Lurie, J. D., Hilibrand, A. S., Tosteson, A. N. A., & Weinstein, J. N. (2005). Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among Spine Patient Outcomes Research Trial (SPORT) participants. *Quality of Life Research, 14(6)*, 1321–1332. https://doi.org/10.1007/s11136-004-5743-2

[45] Ostelo, R. W. J. G., et al. (2008). Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine, 33(1)*, 90–94. https://doi.org/10.1097/BRS.0b013e31815e3a10

[46] Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10(4)*, 407–415. https://doi.org/10.1016/0197-2456(89)90005-6

[47] Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology, 61(2)*, 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

[48] Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: A theoretical model. *Social Science & Medicine, 48(11)*, 1507–1515. https://doi.org/10.1016/S0277-9536(99)00045-3

[49] Fayers, P. M., & Machin, D. (2013). *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes* (2nd ed.). Wiley-Blackwell.

[50] Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

[51] van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54(2)*, 119–135. https://doi.org/10.1016/j.erap.2003.12.004

[52] Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

[53] Klapproth, C.P., Fischer, F., & Rose, M. (2023). Scale agreement, ceiling and floor effects, construct validity, and relative efficiency of the PROPr and EQ-5D-3L in low back pain patients. *Health and Quality of Life Outcomes, 21(1)*, 107. https://doi.org/10.1186/s12955-023-02188-w

[54] Greenhalgh, J., Dalkin, S., Gooding, K., Gibbons, E., Wright, J., Valderas, J. M., & Black, N. (2017). Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. *Health Services and Delivery Research, 5*(2). https://doi.org/10.3310/hsdr05020

[55] Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education, 17*(4), 137–144. https://doi.org/10.1093/bjaed/mkw060

[56] Stover, A. M., Haverman, L., van Oers, H. A., Greenhalgh, J., Potter, C. M., & Santana, M. J. (2015). Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROMs) in routine practice. *Quality of Life Research, 24*(5), 1069–1077. https://doi.org/10.1007/s11136-014-0842-3

[57] Glenwright, B. G., Simmich, J., Cottrell, M., O'Leary, S. P., Sullivan, C., Pole, J. D., & Russell, T. (2023). Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. *Journal of Patient-Reported Outcomes, 7*(1), 13. https://doi.org/10.1186/s41687-023-00554-2

[58] Lans, A., Bales, J. R., Borkhetaria, P., Schwab, J. H., Verlaan, J. J., Rossi, L. P., & Tobert, D. G. (2023). Impact of Health Literacy on Self-Reported Health Outcomes in Spine Patients. *Spine, 48*(7), E87–E93. https://doi.org/10.1097/BRS.0000000000004495

[59] Tuinenburg, A., Determann, D., Quik, E. H., van der Willik, E. M., Hofstra, G., Hallegraeff, J. M., Vriend, I., Warmerdam, L., van Bommel, H. E., Boland, G., & Oude Voshaar, M. A. H. (2025). Evaluating Comprehensibility of 157 Patient-Reported Outcome Measures (PROMs) in the Nationwide Dutch Outcome-Based Healthcare Program: More Attention for Comprehensibility of PROMs is Needed. *The Patient, 18*(1), 65–76. https://doi.org/10.1007/s40271-024-00710-w

[60] Jung, A., Challoumas, D., Pagels, L., Armijo-Olivo, S., Braun, T., & Luedtke, K. (2024). Guidelines for the development and validation of patient-reported outcome measures: a scoping review. *BMJ Evidence-Based Medicine, 29*(6), 363–373. https://doi.org/10.1136/bmjebm-2023-112681

[61] Konopka, J. A., Bloom, D. A., Lawrence, K. W., Oeding, J. F., Schwarzkopf, R., & Lajam, C. M. (2023). Non-English Speakers and Socioeconomic Minorities are Significantly Less Likely to Complete Patient-Reported Outcome Measures for Total Hip and Knee Arthroplasty: Analysis of 16,119 Cases. *The Journal of Arthroplasty, 38*(7 Suppl 2), S69–S77. https://doi.org/10.1016/j.arth.2023.01.005

[62] Rodrigues, R., Silva, P. S., Cunha, M., Vaz, R., & Pereira, P. (2018). Can We Assess the Success of Surgery for Degenerative Spinal Diseases Using Patients' Recall of Their Preoperative Status? *World Neurosurgery, 115*, e768–e773. https://doi.org/10.1016/j.wneu.2018.04.174

[63] Alma, H. J., de Jong, C., Jelusic, D., et al. (2018). Assessing health status over time: impact of recall period and anchor question on the minimal clinically important difference of COPD health status tools. *Health Quality of Life Outcomes, 16*, 130. https://doi.org/10.1186/s12955-018-0950-7

[64] Macchiarola, L., Pirone, M., Grassi, A., Pizza, N., Trisolino, G., Stilli, S., & Zaffagnini, S. (2022). High recall bias in retrospective assessment of the pediatric International Knee Documentation Committee Questionnaire (Pedi-IKDC) in children with knee pathologies. *Knee Surgery, Sports Traumatology, Arthroscopy, 30*(10), 3361–3366. https://doi.org/10.1007/s00167-022-06922-7

[65] Gotlin, M. J., Baron, S., Kingery, M. T., et al. (2019). Recall Bias in Retrospective Assessment of Preoperative Patient Reported Outcomes. *Orthopaedic Journal of Sports Medicine, 7*(7_suppl5). https://doi.org/10.1177/2325967119S00339

[66] Pellisé, F., Vidal, X., Hernández, A., Cedraschi, C., Bagó, J., & Villanueva, C. (2005). Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine, 30*(3), 365–368. https://doi.org/10.1097/01.brs.0000152096.48237.7c

[67] Raphael, K. (1987). Recall bias: A proposal for assessment and control. *International Journal of Epidemiology, 16*(2), 167–170. https://doi.org/10.1093/ije/16.2.167

[68] Surbone, A., & Lowenstein, J. (2003). Exploring asymmetry in the relationship between patients and physicians. *The Journal of Clinical Ethics, 14*(3), 183–188.

[69] Zini, M. L. L., & Banfi, G. (2021). A narrative literature review of bias in collecting patient reported outcomes measures (PROMs). *International Journal of Environmental Research and Public Health, 18*(23), 12445. https://doi.org/10.3390/ijerph182312445

[70] Haverman, L., Luijten, M. A. J., Blackford, A. L., Absolom, K., Basch, E. M., van Rossum, M. A. J., Engelen, V., Grootenhuis, M. A., Velikova, G., & Snyder, C. (2024). Truth and dare: patients dare to tell the truth when using PROMs in clinical practice. *Quality of Life Research, 33*(12), 3299–3307. https://doi.org/10.1007/s11136-024-03772-3

[71] Mannion, A. F., Porchet, F., Kleinstück, F. S., Lattig, F., Jeszenszky, D., Bartanusz, V., Dvorak, J., & Grob, D. (2009). The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. *European Spine Journal, 18* Suppl 3(Suppl 3), 374–379. https://doi.org/10.1007/s00586-009-0931-y

[72] Miekisiak, G., Kollataj, M., Dobrogowski, J., Kloc, W., Libionka, W., Banach, M., Latka, D., Sobolewski, T., Sulewski, A., Nowakowski, A., Kiwic, G., Pala, A., & Potaczek, T. (2013). Cross-cultural adaptation and validation of the Polish version of the core outcome measures index for low back pain. *European Spine Journal, 22*(5), 995–1001. https://doi.org/10.1007/s00586-012-2607-2

[73] World Health Organization. (2024). Portugal: Country profile – European Health Information Gateway. WHO Regional Office for Europe. Retrieved May 31, 2025, from https://gateway.euro.who.int/en/country-profiles/portugal/

[74] Statistics Portugal (Instituto Nacional de Estatística). (2023). Foreign population with legal residence status in Portugal. Retrieved June 1, 2025, from https://dados.gov.pt/en/datasets/populacao-estrangeira-com-estatuto-legal-de-residente-n-o/

[75] Fehlings, M. G., Tetreault, L. A., Riew, K. D., Middleton, J. W., Aarabi, B., Arnold, P. M., Brodke, D. S., Burns, A. S., Carette, S., Chen, R., Chiba, K., Dettori, J. R., Furlan, J. C., Harrop, J. S., Holly, L. T., Kalsi-Ryan, S., Kotter, M., Kwon, B. K., Martin, A. R., Milligan, J., Wang, J. C. (2017). A clinical practice guideline for the management of patients with degenerative cervical myelopathy: Recommendations for patients with mild, moderate, and severe disease and nonmyelopathic patients with evidence of cord compression. *Global Spine Journal, 7*(3 Suppl), 70S–83S. https://doi.org/10.1177/2192568217701914

[76] Smith, E. R., Roldán, M. A., Arvin, B., & Fehlings, M. G. (2021). Prevalence of asymptomatic cervical spinal cord compression: A systematic review and meta-analysis. *Global Spine Journal, 11*(2), 243–252. https://doi.org/10.1177/2192568220934496

[77] Gulati, S., Vangen-Lønne, V., Nygaard, Ø. P., Gulati, A. M., Hammer, T. A., Johansen, T. O., Peul, W. C., Salvesen, Ø. O., & Solberg, T. K. (2021). Surgery for degenerative cervical myelopathy: A nationwide registry-based observational study with patient-reported outcomes. *Neurosurgery, 89*(4), 704–711. https://doi.org/10.1093/neuros/nyab259

[78] Morishita, Y., Buser, Z., D'Oro, A., Shiba, K., & Wang, J. C. (2018). Clinical relationship of degenerative changes between the cervical and lumbar spine. *Asian Spine Journal, 12*(2), 343–348. https://doi.org/10.4184/asj.2018.12.2.343

[79] Neifert, S. N., Martini, M. L., Yuk, F., McNeill, I. T., Caridi, J. M., Steinberger, J., & Oermann, E. K. (2020). Predicting trends in cervical spinal surgery in the United States from 2020 to 2040. *World Neurosurgery, 141*, e175–e181. https://doi.org/10.1016/j.wneu.2020.05.055

[80] Kobayashi, K., Sato, K., Kato, F., Kanemura, T., Yoshihara, H., Sakai, Y., Shinjo, R., Ohara, T., Yagi, H., Matsubara, Y., Ando, K., Nakashima, H., & Imagama, S. (2022). Trends in the numbers of spine surgeries and spine surgeons over the past 15 years. *Nagoya Journal of Medical Science, 84*(1), 155–162. https://doi.org/10.18999/nagjms.84.1.155

[81] Mannion, A. F., Porchet, F., Kleinstück, F. S., Lattig, F., Jeszenszky, D., Bartanusz, V., Dvorak, J., & Grob, D. (2009). The quality of spine surgery from the patient's perspective. Part 1: the Core Outcome Measures Index in clinical practice. *European Spine Journal, 18* Suppl 3(Suppl 3), 367–373. https://doi.org/10.1007/s00586-009-0942-8