



## ORIGINAL ARTICLE

## Development of a Novel Retrospective Instrument to Gauge the General Well-being Index and Lumbar Specific Disability Index Post Intervention in Lumbar Degenerative Disease

Song Ying Chong, MBBS<sup>1</sup>, Mahmoud Danaee, PhD<sup>2</sup> and Dharmendra Ganesan, FRCS<sup>3\*</sup>

<sup>1</sup>Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia

<sup>2</sup>Social and Preventive Medicine, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia

<sup>3</sup>Neurosurgical Unit, Department of Surgery, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia

\*Corresponding author: Dharmendra Ganesan, FRCS (Neuro.Surg), Neurosurgical Unit, Department of Surgery, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia



### Abstract

The clinical outcome data for patients who had a particular intervention for the lumbar degenerative disease is based on changes in symptomatology. This functional data had been acquired in a prospective manner using standard tools e.g., ODI, RMQ, SF-36 etc. Such data may be lacking in many facilities of developing countries due to various reasons. We are proposing a retrospective comparative questionnaire to obtain similar outcome data after intervention for patients with lumbar degenerative spinal disease. The process of content validation and reliability of the questionnaires known as the UM (University of Malaya) general well-being index post-intervention for spinal disorders and UM post-intervention lumbar specific spinal disability index is presented. This questionnaire is simple but encompasses key functional outcomes. There is a summative score that would indicate the overall outcome. It can be conducted efficiently in person or via the telephone. It would require further validation in clinical use for patients with lumbar degenerative disease.

### Keywords

Lumbar degenerative spinal disease, Outcome, Questionnaire, Content validation, Reliability, Quality of life (QOL)

### Abbreviations

ADL: Activities of Daily Living; CVI: Content Validity Index; EFA: Exploratory Factor Analysis (EFA); ODI: Oswestry Disability Index; PA: Parallel Analysis; RMQ: Roland-Morris Questionnaire; QOL: Quality of Life; SF-36: Short Form-36 Questionnaire.

### Introduction

Research concerning degenerative spinal disorders relies on the patient's perspective on their quality-of-life (QOL) to assess the effect of the treatment [1]. There were multiple questionnaires available in the field which evaluates their spine-related disability index, including the Oswestry Disability Index [2], Roland-Morris Classification [3] and Quebec Classification Score [4]. While other questionnaires were used to assess the patient's quality of life, namely Short Form-36 (SF-36) [5] and EQ5D questionnaires [6]. These are comprehensive questionnaires that serve as a good tool to gauge patients' specific and general functional state at one point in time taking into account the symptoms and their effects on daily activities and quality of life. If this assessment is done at the various timelines, it can be used to assess the response to a specific therapeutic modality.

There are various outcome instruments available and the selection of the most appropriate instrument would be determined by the researcher [7]. The commonly used questionnaire Oswestry Disability Index [2], Roland-Morris Classification [3], Short Form-36 (SF-36) [5], need data recorded in a prospective manner at various points in time to assess the progress of the condition or response to treatment. This would

require infrastructure and manpower to conduct such research as it is time-consuming. In many parts of the developing world, this sort of facility and support is lacking; even if it were available sometimes in a busy clinical practice such evaluation could not be conducted on every patient.

Lumbar degenerative disease is a common disorder. There is a cohort of patients who would benefit from surgical intervention. The surgical options for degenerative spinal disorders are varied; ranging from simple neural decompression using microscope or endoscope; to more intricate approaches involving fusion surgery. The surgical options that are advocated by the surgeon are based on their training, experience, and locality. In the current times, the literature pertaining to this subject is predominantly from developed countries. The surgical strategies employed in the developed world to manage lumbar degenerative disease may differ from the developing or less developed countries. There is a paucity of literature on the techniques used and the outcomes regarding degenerative spinal disease from the developing world. Because of the limited resources, these countries tend to practice a more basic neural decompression surgery rather than fusion surgery; however, this outcome data is not widely available because of limitation of obtaining prospective data in these countries due to various reasons including lack of funding, heavy clinic load, logistics and manpower. Therefore, these sets of data do not get into mainstream literature but are possibly presented only in local meetings in a form of the surgeons' experience. In an effort to capture this basic outcome data from this cohort of patients, we embarked on devising a truncated questionnaire that permits for a broader retrospective comparative evaluation of general well-being and spine specific disabilities for lumbar degenerative spinal disorder patients who have had an intervention. The outcome measures would permit the researcher to appraise the benefit of a particular intervention in terms of functional domains.

The questionnaire has two large domains of general well-being and lumbar specific spinal disabilities index. Each domain has five questions, and the responses are objective with either one of the three to four responses. The choice of questions in each domain is general (pain intensity, personal care, sleeping, social life engagement/recreation and work) and specific (walking, sitting, change in position, bladder dysfunction, lower limb numbness). These parameters were chosen in an attempt to reflect a more holistic appraisal of the functional state despite it being a short questionnaire. This assessment can be done retrospectively either in person or through the telephone which brings out the advantages of greater response rate, relatively inexpensive but quick data collection and the ability for interviewers to probe for incomplete or ambiguous answers [8]. This method is also relevant in the current

pandemic. We are presenting our assessment tool and the validation process.

Content validation certifies whether an instrument measures what is intended and verifies whether the items properly reflect the content domain of interest and whether the scale dimensions are consistent with each proposed item and provide clarity to respondents [9,10]. Content validity aspects include the relevance of the item and its clarity, which are determined by an evaluation of the instrument items by a group of experts with previous experience or currently recognized in the field of study, namely practicing surgeons, physicians or experts [11]. The present study aimed to validate the content of a developed instrument to assess the general well-being and quality of life of individuals with the lumbar degenerative disease who underwent lumbar spine surgeries (intervention).

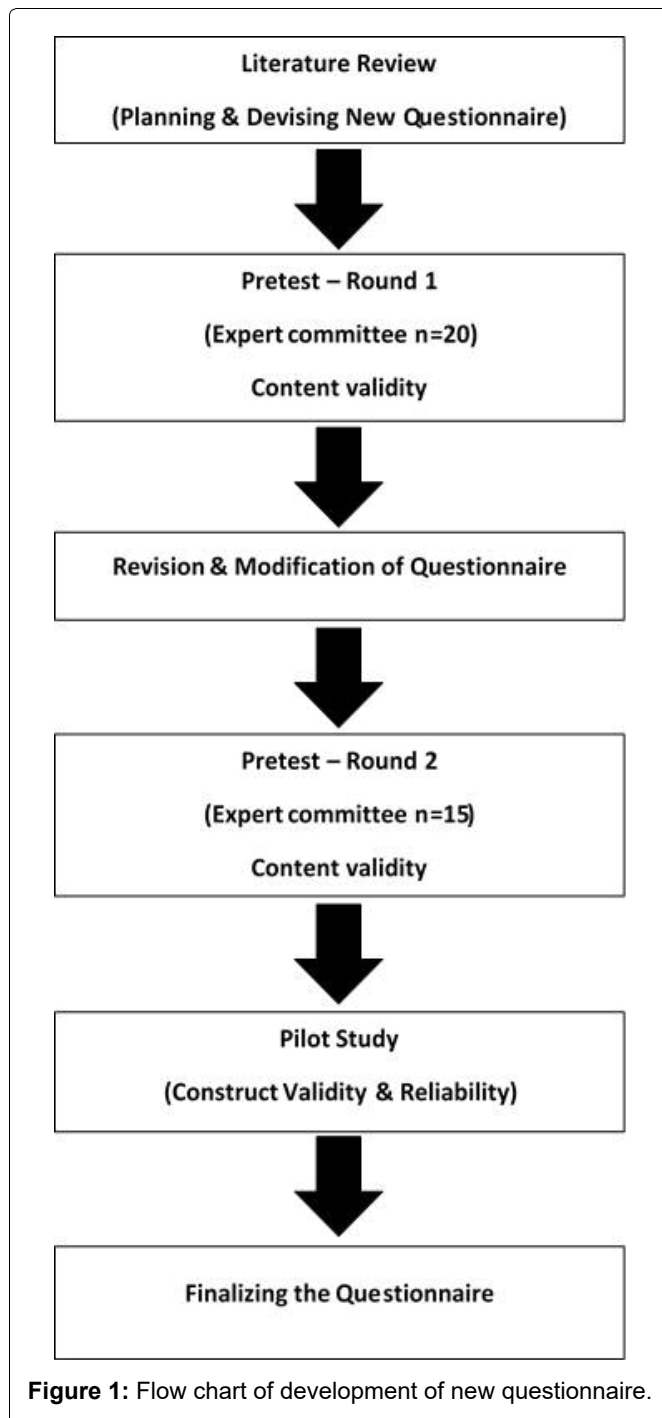
## Method

The stages for the development and validation of an instrument encompass four distinct phases: Planning, construction, quantitative analysis, and validation, and lastly pilot study among patients [11]. The instrument planning and construction phases consisted of an extensive literature review and expert opinions. Content validation and the pilot study are described in the present study. This project was approved according to protocol no. 2020819-8994 of the Medical Research Ethics Committee, University Malaya Medical Centre, Kuala Lumpur, Malaysia.

The content validation (phase I) was done in two rounds. In the first round, the instrument was initially subjected to content validation by a committee of twenty medical experts with an interest in the field of degenerative spinal disease. The content validation questionnaire was developed in English, it was evaluated for the relevance and clarity of each item developed. The second round of content validation was repeated among fifteen experts after the initial responses and comments were considered and relevant changes made to fine-tune the instruments. Once the content validations by the experts were concluded a pilot study was conducted to evaluate the construct validity. [Figure 1](#) demonstrates a flow chart in the development of a new questionnaire.

## Participants

**Phase I (Pre-test):** Based on the protocol described by Polit, et al. [12], we choose twenty medical practitioners including neurosurgeons, orthopaedic spine surgeons, rheumatologists, neuropsychologist, and rehabilitation physician with experience in the care of patients with the lumbar degenerative disease to participate. These included both local and foreign medical practitioners. The length of experience of the experts ranged from 8 to 25 years. In the first round of content validation, twenty experts with seventeen local Malaysian experts and three international experts responded within a



specified period while fifteen experts from a similar cohort participated in the second round.

**Phase II (Pilot study):** A pilot study was conducted from a pool of patients who underwent lumbar spine surgeries for degenerative spine diseases in a tertiary hospital in Kuala Lumpur from the period of January 2008 to December 2019. As such, a total of 104 patients were successfully contacted and interviewed via telephone within a period of 2 weeks. The minimum sample size for EFA, according to Gorsuch [13], should be at least 100, therefore the collected data ( $n = 104$ ) was adequate to calculate factor loadings for each question.

### Data analysis

**Phase I (Pre-test):** The content validity of the developed instrument was statistically analyzed by a Content Validity Index (CVI). To calculate the CVI, each item was ranked on a four-point scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant). For each item, the CVI was calculated as the number of experts who provided a rank of 3 or 4, divided by the total number of experts. The Kappa Modified Coefficient was used to determine the degree of relevance agreement of the CVI and was calculated according to protocols set by Polit, et al. Twenty experts participated in the first phase of content validation, the acceptable CVI value for each item ranged from 1.00 to 0.70, and the Kappa Modified value from 1.00 to 0.68. In the second phase, the items considered as acceptable were those that had a CVI ranging from 1.00 to 0.80 and a Modified Kappa value ranging from 1.00 to 0.80 because fifteen experts participated in these phases. Because the CVI considered the number of experts consulted for each phase, the difference in the number of experts in the different content validation phases of the present study did not influence the statistical results.

**Phase II (Pilot study):** To do the construct validity of the questionnaire, the pilot study data were analyzed first through Parallel Analysis. Parallel Analysis is a Monte

**Table 1:** The first phase of content validation of the general disability domain.

Domain		Items	Relevance		Clarity		Results
			CVI*	Kappa	CVI*	Kappa	
General Disability	1	Pain Intensity	0.95	0.95	0.7	0.688	Corrected
	2	Personal Care (ADL)	0.95	0.95	0.9	0.9	Validated
	3	Sleeping	0.75	0.746	0.75	0.746	Corrected
	4	Social Life Engagement & Recreation	0.9	0.9	0.75	0.746	Corrected
	5	Work	0.95	0.95	0.85	0.85	Validated
Lumbar Specific Disability	6	Walking	0.95	0.95	0.9	0.9	Validated
	7	Sitting	0.9	0.9	0.9	0.9	Validated
	8	Change in Position	0.75	0.75	0.85	0.85	Corrected
	9	Bladder Dysfunction	1	1	0.75	0.75	Corrected
	10	Lower Limb Numbness	0.9	0.9	0.95	0.95	Corrected

\*CVI: Content validity index.

Carlo simulation method that helps scholars in defining the number of factors to retain in Principal Component and Exploratory Factor Analysis. This technique delivers a superior difference to other techniques that are generally used for the same purpose, such as the Scree test or the Kaiser's eigenvalue-greater-than-one rule. In this study the parallel analysis was done based on this method using online software for PA ([www.statatoddo.com](http://www.statatoddo.com)). Parallel analysis (PA) is a method used to define the number of factors in factor analysis. There are several factors that affect the results of a PA: the choice of the eigenvalue percentile, the strength of the factor loadings, the number of variables, and the sample size of the study. Followed by parallel analysis, Exploratory Factor Analysis (EFA) was applied to evaluate the construct validity and to determine the contribution of each item (loading factor) to related components (factor). The EFA was done using Varimax rotation to determine the factor structure among 10 items related to the scale. All statistical test was done using SPSS Ver 25.

## Results

### Phase I

**The first round of content validation:** In the first round of content validation, the domains were considered satisfactory (CVI between 0.70 and 1.00) for evaluating the general disability and the region-specific domains. All items in the General Characteristics Domain obtained satisfactory scores ([Table 1](#)) and hence all the items in this domain were maintained in the second version of the instrument. Item 3 which assessed sleeping although achieved acceptable scores in relevancy (CVI = 0.75, Kappa = 0.746) and clarity (CVI = 0.70, Kappa = 0.746) was modified to ascertain the sleeping problems was due to the spinal problems instead of other issues as pointed out by one of the experts. The items that assessed pain intensity and social life engagement & recreation also had their wordings changed for improved understanding, for example in the pain intensity section, the options of "improved,

maintained and worsened" were changed to "reduced, persistent and worsened". Items 6-10 presents the common scores of the lumbar-specific items. All items achieved a CVI of more than 0.85 except for relevancies in item 8 which assesses the change in positions of patients (CVI = 0.75) and clarity in item 9 accessing patient's bladder dysfunction (CVI = 0.75). To keep the questions more focused and maintain clarity to obtain a consistent understanding of the question either by reading or listening to the question by the respondent, we added the phrase "After the spinal intervention, ..." to each of the questions. Wordings were also improved while some grammatical errors were pointed out by the experts to keep the questions all in the present tense.

**The second round of content validation:** [Table 2](#) represents the content validation of the general disability domain and lumbar-specific domain in the second round of the content validation process. The second version of the instrument was resubmitted to a similar expert committee for reevaluation of all items. 15 experts responded and participated in the second phase of content validation. All items managed to achieve a CVI of > 0.85 for both relevancy and clarity. Hence, all items were successfully validated after appropriate corrections were made prior to the second phase of validation.

### Parallel analysis

In this section, the results of the parallel analysis are presented for the adapted instrument including 10 items. The parallel analysis ([Figure 2](#)) indicated that only the eigenvalue of the first extracted factor was more than the eigenvalue of simulated data ( $\lambda = 1.646$ ) which suggest that a single factor can be extracted from the data and therefore in the next step for exploratory factor analysis the number of extracted components is considered as 1 factor.

### EFA of impact scale

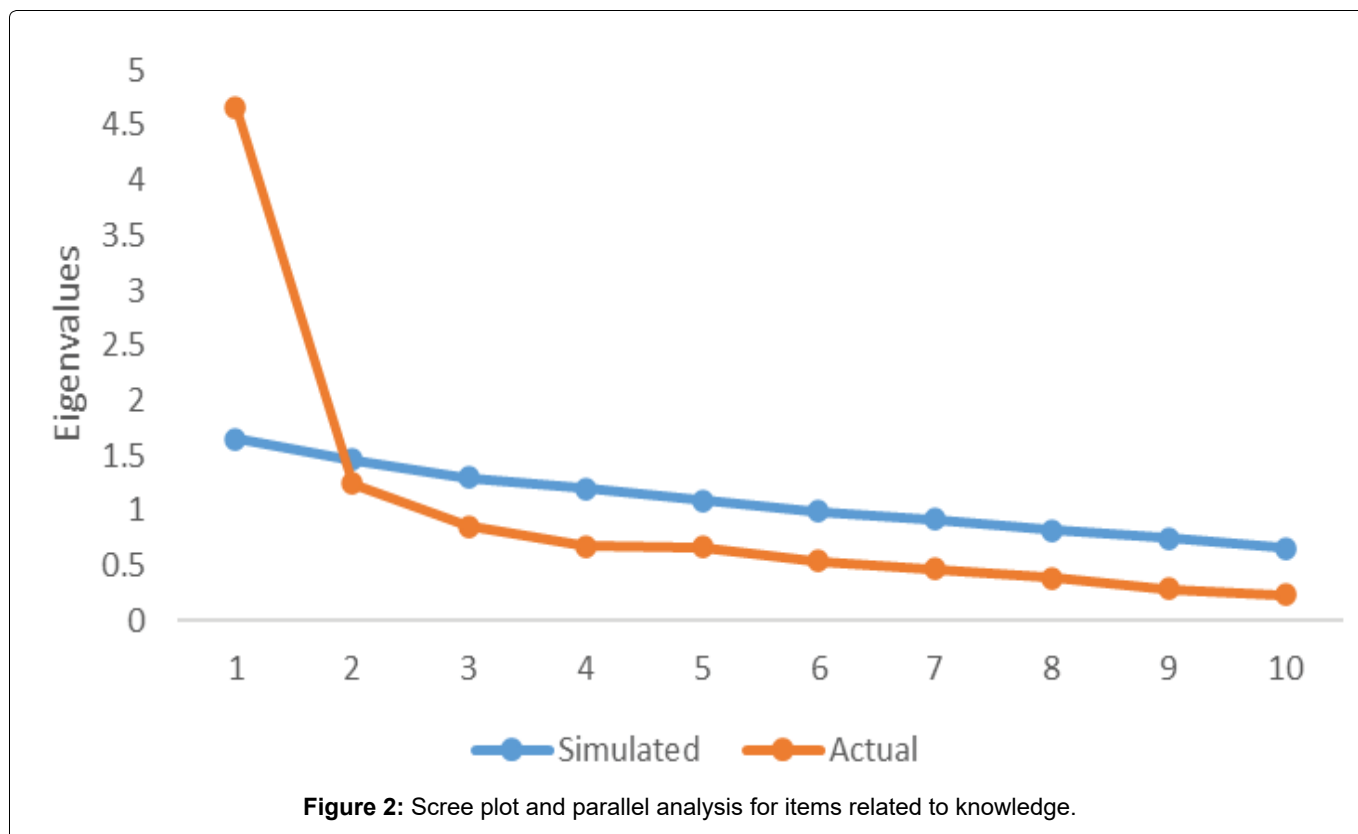
Exploratory Factor Analysis was using Varimax rotation was applied to determine the factor structure

**Table 2:** The second phase of content validation of the general disability domain.

Domain	Items	Relevance		Clarity		Results	
		CVI*	Kappa	CVI*	Kappa		
General Disability	1	Pain Intensity	0.933	0.933	1	1	Validated
	2	Personal Care (ADL)	1	1	1	1	Validated
	3	Sleeping	0.933	0.933	0.867	0.866	Validated
	4	Social Life Engagement & Recreation	1	1	1	1	Validated
	5	Work	1	1	1	1	Validated
Lumbar Specific Disability	6	Walking	1	1	0.933	0.933	Validated
	7	Sitting	0.933	0.933	1	1	Validated
	8	Change in Position	1	1	1	1	Validated
	9	Bladder Dysfunction	0.933	0.933	1	1	Validated
	10	Lower Limb Numbness	0.933	0.933	0.867	0.867	Validated

\*CVI: Content validity index.





**Table 3:** Factor loadings based principal component analysis with varimax rotation for 10 items related to general well-being scale.

Items	Factor loading
Q5	0.844
Q6	0.796
Q1	0.776
Q8	0.768
Q4	0.764
Q2	0.700
Q3	0.660
Q7	0.552
Q9*	0.396
Q10*	0.390

**Extraction Method:** Principal component analysis.

**Rotation Method:** Varimax with Kaiser normalization.

among 10 items related to the general well-being scale. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.863, above the suggested value of 0.6, and Bartlett's test of sphericity was significant ( $\chi^2_{(45)} = 420.581, p < 0.05$ ).

In the current study, all initial communalities were above the threshold. The results of EFA on all 10 items were forced into one component based on the results of the parallel analysis. The results showed that this component explained 46.6 % of the variance. According to these results, as shown in Table 3, all items showed a loading factor above 0.5 except item 9 and item 10 had a loading less than 0.5.

**Table 4:** Exploratory factor analysis (EFA) based on time (Loading Factors).

	Before 2010	2011-2013	2014-2016	After 2017
Q1	0.889	0.687	0.701	0.833
Q2	0.932	0.797	0.702	0.642
Q3	0.739	0.888	0.761	0.45
Q4	0.822	0.797	0.775	0.764
Q5	0.844	0.812	0.86	0.822
Q6	0.724	0.931	0.839	0.745
Q7	0.740	0.354	0.539	0.633
Q8	0.667	0.793	0.894	0.758
Q9	0.932	0.407	0.400	0.03
Q10	0.211	0.318	0.496	0.446

In order to investigate the effect of time on the outcomes of construct validity, the sample was divided into four categories and EFA was performed on each subset of data (Table 4) because the current study is based on a retrospective study that took place from 2008 to 2020. The results revealed that, with the exception of Q9 and Q10, which had low loading factors in the pooled sample, the remainder of the items had acceptable loading factors across different subgroups, indicating that the passage of time had no effect on the results of the EFA.

## Discussion

Lumbar degenerative spinal disease is a "wear and tear" condition that affects the more mobile parts of the spine mainly the cervical and lumbar region [14]. Most patients' manifest symptoms of pain or numbness

as their primary problem. The threshold to offer surgical intervention also hinges on the severity of the symptoms leading to functional disability in executing the activities of daily living as well as the quality of life in terms of vocation and recreation. Therefore, the success of a particular intervention rendered to a patient with the degenerative spinal disease can be gauged by assessing the resolution or improvement of the presenting symptoms [15], where the ideal outcome of treatment is a return to the normal or usual quality of life for a given age and medical condition [16]. This gives a good indicator of the functional benefit that the patient had derived from a particular intervention/treatment within his own environment. Therefore, in cases of degenerative lumbar spinal disease, the symptomatology together with functional disability play a key part in making the management decisions as well as assessment of outcomes of these patients before and after the intervention.

Wilson and Cleary proposed a taxonomy for different measures of health outcome, which divides these outcomes into five levels: Biological and physiological factors, symptoms, functioning, general health perceptions, and overall quality of life [17]. Consequently, the lumbar spine specific and general well-being questionnaires in various forms have been used to this end [2-4]. However, the key features of these questionnaires are they are conducted in a prospective manner indicating the symptoms and functional disability at that one point in time. Some of these questionnaires are lengthy having multiple stems for one particular symptom that is being assessed eg. Oswestry Disability Index [2], by doing so it is able to tease the finer details of a symptom or activity. Though it is a more detailed assessment tool acquiring these data accurately from the subjects needs patients cooperation and compliance. These questionnaires have undergone vigorous validation and are a standard means of furnishing outcome data for degenerative spinal disorders in many works of literature.

Given the busy service and lack of research assistants, prospective functional data are not routinely collected for most of the cases of degenerative spinal disease that sought treatment in our facility. Based on the clinical follow-up, it is apparent that most of the patients who have undergone surgical intervention seem to have improved in functional ability. We wanted to study our surgical outcomes in a more scientific manner, but we realized that we lacked the prospective data. This initiated us to devise an assessment tool to perform a retrospective comparative assessment in a more standardized manner to allow a more systematic assessment of the outcome. The retrospective assessment tool will allow a comparison of the perceived degree of improvement in symptoms before and after the intervention. This comparison can be done at the various timeline from the intervention.

However, the further the timeline is from the index surgery there is the possibility that the patient may not be able to make a reliable comparison. To minimize this sort of bias, we keep the questions more functional in terms of their overall experience in a particular item of assessment rather than asking the patient to score the specific symptoms in a retrospective manner. Scoring symptoms retrospectively may be fraught with errors. However, evaluating the patient's experience of change in various functional items after an intervention in a broader sense rather than specific scoring would minimize the errors in terms of recall biasness. Without a retrospective instrument, we cannot measure the result more objectively in a standardized manner due to the lack of prospective functional data.

The stark limitation of these existing questionnaires eg. ODI, RMQ, SF-36, etc. is the fact that the data must be collected in a prospective fashion at two different time points eg. before and after the intervention to make the comparison of the outcome after a particular intervention. It would be a more sensitive and specific way of presenting the outcome data, but to obtain such data would need resources and patient commitment to answer these detailed questionnaires. In order to achieve, a similar endpoint of outcome assessment after a particular intervention in a patient with degenerative spinal disease; we have validated a single assessment tool using a comparative questionnaire to gauge the difference that the patient has experienced after the intervention. The questionnaire looks at a few common functional parameters to be a surrogate of the overall functional ability of the patient. The questions do not interrogate details of the symptom or activities like the Oswestry Disability Index (ODI) because the answers expected are more of a reflection of the difference that the patient experiences after the intervention. Therefore, the options within each parameter are kept broader (Appendix 1) but distinct, nevertheless. Despite losing on the finer symptomatology and functional analysis, this assessment tool is more usable and able to derive reasonable outcome assessments to indicate the usefulness of certain interventions in patients with lumbar degenerative disease particularly in resource-limited countries.

The questionnaire underwent content validation and reliability test. Factor loadings are correlation coefficients between observed variables and latent common factors [18]. Factor loadings are part of the outcome from factor analysis, which serves as a data reduction method designed to explain the correlations between observed variables using a smaller number of factors. A factor loading that falls outside of the interval bounded by ( $\pm$  cutoff value) is large and is thus retained. On the other hand, a factor loading that does not meet the criterion indicates that the corresponding observed variable should not load on the corresponding common factor.

The general well-being domains that we focused on were pain intensity, personal care, sleeping, social life engagement/recreation and work. The first question was pertaining to pain intensity. We kept this question broad to encompass overall pain and not specific to the low back pain and lower limb pain. We are assessing the overall pain syndrome related to the degenerative spinal disorder and hope that by tackling the primary pain source, the other secondary pain related issues will resolve or diminish. However, there were limitations to this question, as addressed by the experts, where patients would sometimes confuse non-spine related pain at the same time and resulting in differing expected results. However, this question has a factor loading of 0.776, which makes it a strong indicator that this question is highly relevant and important to access the patient's general well-being outcome. The second question included in the questionnaire was about personal care and activity of daily living (ADL). Options with concepts of being fully ADL independent, partially dependent and total dependent were being implemented in this section. Like the first question, we appreciate there may be other factors affecting the independence to daily living apart from the spine issue. This question has a factor loading of 0.700, which also makes up a relatively strong contributor to a patient's quality of life.

The lumbar spine specific domains that we focused on were walking, sitting, changing position, bladder dysfunction and lower limb numbness as these functions had specific relation to the lumbar spine. For example, in the first question which focuses on walking, we emphasized any presence of back or leg pain and if it causes any restriction in walking as a result of the back pain. This question has an acceptable factor loading of 0.796. The only 2 items which had a factor loading of  $< 0.400$  were domains of bladder dysfunction (0.396) and lower limb numbness (0.390). This could be explained as the incidence of bladder dysfunction because of degenerative spine disease was relatively uncommon even before surgery in our pool of patients (14.4%), hence making this question irrelevant in most of the patients in the pool. Meanwhile, in the domain of lower limb numbness, patients may perceive numbness in different ways; while some patients felt the residual numbness was acceptable and negligible while some patients still felt the numbness in the lower limbs affecting their daily activities. Most papers considered loading factors greater than 0.5 and any lower than that should be excluded and removed from the questionnaire. However, according to [18-21], the cutoff value is arbitrarily selected depending on the field of study, and ( $\pm 0.4$ ) seems to be preferred by many researchers. Hence, as the values of questions 9 & 10 were close to the cutoff point of 0.4, it is not being excluded in the current study. However, further evaluation should be made in future studies to determine its value and to

decide on whether these 2 items should be excluded in the future.

The date of the intervention and the date of the assessment is recorded in the questionnaire to determine the timeline. The same instrument can be used to record outcome measures at various timelines. Hence, analyzing the benefit of that index intervention at various points in time. Bearing in mind, the confounding factors that would affect the outcome apart from the intervention would have a similar effect on the subject's assessment of symptomatology in both the prospective and retrospective questionnaire.

In order to obtain an overall impression of the outcome after an intervention, the outcome for each item in the questionnaire has been denoted a score. We are proposing a summative score for the general well-being index and lumbar specific disability index respectively. A scale is used where the general well-being index can be scored between 5 to 16 and the lumbar specific disability index scored between 5 to 15. These scores are summated and then grouped randomly into four categories to indicate the overall outcome (excellent, good, moderate, poor) (Appendix 2). Once in clinical use, further clinical data will be employed to evaluate the discriminative and predictive validity as well as optimize the cut-off point for the scale.

There will be certain limitations of a retrospective questionnaire to assess functional outcomes. One of the key elements is the ability to remember the severity of the symptomatology at presentation. The risk of this biasness has been alleviated by focusing on the experiential change as perceived by the subject of the symptoms in functional terms rather than grading the severity of the symptom at presentation in a retrospective manner. The analysis on the effect of time on the outcomes of construct validity indicated that the passage of time had no effect on the results of the EFA (Table 4). This is a novel instrument to permit a more standardized means of measuring outcomes for the cohort of lumbar degenerative disease patients who do not have prospective clinical data. By having such an instrument, more developing countries would be able to describe the outcomes in reference to the time of intervention of various sorts of treatments that are performed for patients with lumbar degenerative spinal disease.

## Conclusion

The content validation of the UM general well-being index post-intervention for spinal disorders and UM post-intervention lumbar specific spinal disability index shows it is a valid and reliable instrument to give an impression of the outcome after intervention in a patient with degenerative lumbar spinal disease. The functional parameters that were gauged were relevant to the disease condition. This tool will afford

a standardized means of performing the retrospective comparative assessment of the effectiveness of a particular intervention. This is a valuable tool to describe outcome measures in a cohort of patients who do not have prospective pre-intervention data.

## References

1. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, et al. (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. *Spine J* 8: 968-974.
2. Fairbank J, Couper J, Davies J, O'Brien J (1980) Oswestry low back pain disability questionnaire. *Physiotherapy* 66: 271-273.
3. Roland MO (1983) The natural history of back pain. *Practitioner* 227: 1119-1122.
4. Kopec JA, Esdaile JM, Abrahamowicz M, L Abenhaim, S Wood-Dauphinee, et al. (1995) The Quebec back pain disability scale. Measurement properties. *Spine (Phila Pa 1976)* 20: 341-352.
5. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (Sf-36). I. Conceptual framework and item selection. *Med Care* 30: 473-483.
6. Group TE (1990) EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 16: 199-208.
7. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, et al. (2009) Outcome instruments: Rationale for their use. *J Bone Joint Surg Am* 91: 41-49.
8. Jackowski D, Guyatt G (2003) A guide to health measurement. *Clin Orthop Relat Res*, 80-89.
9. Benson J, Clark F (1982) A guide for instrument development and validation. *Am J Occup Ther* 36: 789-800.
10. Davis AE (1996) Instrument development: Getting started. *J Neurosci Nurs* 28: 204-207.
11. Natalio MA, Faria CDCM, Teixeira-Salmela LF, Michaelsen SM (2014) Content validation of a clinical assessment instrument for stair ascent and descent in individuals with hemiparesis. *Braz J Phys Ther* 18: 353-363.
12. Polit DF, Beck CT, Owen SV (2007) Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Res Nurs Health* 30: 459-467.
13. Gorsuch RL (1983) Factor analysis. (2<sup>nd</sup> edn), Hillsdale, NJ.
14. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Rothman SLG, et al. (2014) Lumbar disc nomenclature: Version 2.0: Recommendations of the combined task forces of the North American Spine Society, the American Society Of Spine Radiology, and the American Society of Neuroradiology. *Spine (Phila Pa 1976)* 39: E1448-E1465.
15. Ware JE (1993) Measuring patients' views: The optimum outcome measure. *BMJ* 306: 1429-1430.
16. Silver GA (1990) Paul Anthony Lembecke, MD, MPH: A pioneer in medical care evaluation. *Am J Public Health* 80: 342-348.
17. Wilson IB, Cleary PD (1995) Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA* 273: 59-65.
18. Michalos AC (2014) Encyclopedia of quality of life and well-being research. *CMAJ* 141: 1077-1079.
19. Guadagnoli E, Velicer WF (1988) Relation of sample size to the stability of component patterns. *Psychol Bull* 103: 265-275.
20. Pituch KA, Stevens JP (2016) Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS.
21. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60: 34-42.