Validation of two disease-specific quality of life questionnaires comparing a partial credit model and a latent class model

Sara Viviani, Alessio Farcomeni, Gian Luca Di Tanna, Anna Rita Vestri

Department of Experimental Medicine
Sapienza - University of Rome

Corresponding Author:
Dr. Sara Viviani
sara.viviani@uniroma1.it

Summary
Peripheral artery disease (PAD) is a condition in which fatty deposits build up in the inner linings of the artery walls. These blockages restrict blood circulation, mainly in arteries leading to the kidneys, stomach, arms, legs and feet. The purpose of this study is to find the most appropriate model for the validation of two disease-specific quality of life (QOL) questionnaires for patients with chronic limb ischemia: PAQ (Peripheral Artery Questionnaire) and VascuQoL (Vascular Quality of Life questionnaire). Both the questionnaires are often used in clinical trials, but an analysis of their validity is needed, in order to obtain reliable Quality of Life's measures which are important to evaluate the impact of Peripheral Arterial Disease and eventually to inform decision making models. We compared a partial credit model and a latent class model, which highlighted different aspects of the questionnaires' configuration.

KEY WORDS: *Paq, Vascuqol, quality of life, partial credit model, latent class model, peripheral arterial disease.*

1. Introduction

Quality of life (QOL) is a multidimensional concept which may be concerned with a disease and discomfort, but also with the overall sense of wellness, only in part related to the clinical health status.
It is fundamental for patients with peripheral arterial disease (PAD) to evaluate whether the therapy improves not only health status, but as well QOL. Research so far focuses on stability of symptoms and physical abilities, such as the ability to walk, but significant improvements can be obtained for quality of life. Many surveys (see for istance [1]) prove that QOL in PAD is substantially poor, mainly because of depression, observed in about 45% of patients [2]. So it is necessary to measure this aspect, that includes physical but also social and emotional outcomes (such as in [3] and [4]).
PAQ (Peripheral Artery Questionnaire) and VascuQoL (Vascular Quality of Life questionnaire) are two questionnaires often used for the evaluation of Quality of Life's in patients with Peripheral Arterial Disease (PAD). PAQ is a 20-item questionnaire that quantifies patients' physical limitations, symptoms, social function, treatment satisfaction and overall quality of life; VascuQoL is a 25-item questionnaire subdivided into five domains: pain (4 items), symptoms (4 items), activities (8 items), social (2 items) and emotional (7 items).
We administered the questionnaires to the same sample of 100 patients with PAD, enrolled in S.Giovanni Addolorata Day Hospital in Rome and subjected to face-to-face interviews.
Other validation analysis have been just made, such as in [5] for PAQ and [6] and [7] for VascuQoL with smaller sample sizes (respectively 44, 39 and 62 subjects) and standard techniques.
Our approach was based on the partial credit model and the latent class model.
The first one is a classical logistic model which depends on two parameters: the individual's ability to response and the item's difficulty. The ability's pa-

rameter represents the capacity of a single subject to be well described by the scale, while the difficulty's parameter can be interpreted as the importance given by the individuals to a single item. Through the interaction between these two parameters a latent, measurable variable can be assumed to exist under data. This aspect allowed us to construct a quantitative variable which in this context can be seen as a measure of Quality of Life and that characterized each individual. An exploratory analysis based on correspondence analysis confirmed this assumption.
Another approach we used as means of comparison is a simultaneous modeling (through PCM) and clustering, which can be achieved assuming a Latent Class Model (see [8]). The basic idea underlying a latent class (LC) analysis is that some of the parameters of a postulated statistical model differ across unobserved subgroups. These subgroups form the categories of a discrete latent variable.
The rest of the paper is as follows: in Section 2 we describe the data collection and the questionnaires used. In Section 3 we illustrate the methodologies employed and their formal properties. In Section 4 we discuss the results obtained and their implications. The principal conclusions are shown in Section 5.

2. Interviews

The study was carried out on 100 patients afferent to the Angiology Division Day Hospital in S. Giovanni Hospital in Rome and recruited during the period between 2007 and 2008. The patients were subjected to face-to-face interviews using two questionnaires: PAQ and Vascuqol.
The Peripheral Artery Questionnaire (PAQ) is a 20-item questionnaire developed to quantify patients physical limitations, symptoms (frequency, severity, and recent change over time), social function, treatment satisfaction, and overall quality of life. The Vascular Quality of Life questionnaire (VascuQoL) is composed of 25 items subdivided into five domains: pain (4 items), symptoms (4 items), activities (8 items), social (2 items) and emotional (7 items).
The sample is composed of subjects aged between 20 and 95, with a prevalence of individuals between 60 and 69 (58%), substantially equidistibuited by sex (53% female).
The stage of the disease constitutes a relevant confounding factor, the study being transversal, but we have not any direct information about it. Nevertheless some items can provide guidance on the intensity of symptoms: according to the third item of PAQ, 52% of individuals do not have any limitations in walking on a smooth surface; furthermore 54% did not feel pain in calves or legs more than once a week during the previous month (fourth item of PAQ). This leads us to deduce that about one half of the sample is in the early stage of the disease.
Most of the individuals we considered presents comorbidities or complications. The most frequent one is the hypertension (61%), followed by claudication, diabetes, Raynaud's phenomenon, hypercholesterolemia and gangrene.
As in all chronic diseases, PAD produces a relevant psychological impact on patients principally due to the limitations produced. According to the twelfth item of PAQ, 40% of patients felt 'always' or 'often' depressed in the previous month.

3. Methods

A preliminary analysis was applied in order to evaluate the internal consistency and the reliability of the two scales according to this particular sample.
The measures employed were the common correlation coefficients used in this field: Cronbach's Alpha and the Split Half proposed in [9]. The Cronbach's Alpha is given by

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_l \sigma_l^2}{\sigma_{sum}^2}\right) \tag{1}$$

where n is the number of variables in the scale, σ_l^2 the variance of the generic variable l and σ_{sum}^2 the variance of the sum variable. α can assume values between 0 and 1. The Split Half is the Cronbach's Alpha computed for the first and the second part of the questionnaire, measuring the balance of the scale according to the reliability. Through this indexes an overall measurement of internal consistency and reliability is given.
Since Quality of Life is a multidimensional concept, an exploratory analysis based on correspondence analysis was then computed to highlight the dimensions measured. This analysis was then used to critically discuss the possibility of the application of the partial credit model, a fundamental assumption for its application being the one-dimensionality of the latent varia-

ble. In detail we assumed that the scales provided a measurement of a single latent variable influenced by both the ability of the individuals and the difficulty of the items. The outcome of the CA was then used for clustering in order to provide further evidence for data structure.

3.1 The partial credit model

The Rasch Model introduced in [10] is a logistic model based on the assumption that the probability of response to an item by any individual is a function of items difficulty and individuals ability. The mathematical relationship between the difficulty and ability parameters is given by the following theorem:

Theorem 3.1. *Having observed a matrix of responses to J dichotomous items by I subjects, assuming: one-dimensionality, monotone ICC (see below), lack of guessing, local independence of the items, sufficiency of raw scores; the Item Characteristic Curves can be expressed as (**Dichotomous Rasch Model**):*

$$g(\theta_i) = P(X_{ij} = x_{ij}|\theta_i, \beta_j) = \frac{exp[x_{ij}(\theta_i - \beta_j)]}{1 + exp[x_{ij}(\theta_i - \beta_j)]} \quad (2)$$

where X_{ij} is either 0 or 1, and represents the response to item j by individual i; θ_i is the parameter representing the ability of the individual i to answer and β_j the difficulty of the item j.

The extension to arbitrary categorical, i.e. not dichotomous, items does not conceptually modify the model, but introduces more complexity. The general formulation we used was given by the **Partial Credit Model**:

$$P_{ijk} = P(X_{ij} = x_{ij}|\theta_i, \beta_{jk}) = \frac{exp[\sum_{k=0}^{x_{ij}}(\theta_i - \beta_{jk})]}{\sum_{h=0}^{H} exp[\sum_{k=0}^{h}(\theta_i - \beta_{jk})]} \quad (3)$$

with $x_{ij} = 0, 1, \ldots, h, \ldots, Hj$.

For rating-scale data, the model provides not only an estimate of the difficulty, but also a series of threshold, i.e. the level at which the likelihood of not observing a given response category (below the threshold) turns the likelihood of observing that category (above the threshold). The threshold values are estimated once for all items, and the estimation is based on the proportion of the actual responses to the total possible responses.

The measurement resulting from the PCM has the properties of an interval scale, i.e. allows equal intervals between two graduations. Furthermore it is linear (on the logit scale) and quantitative.

Some measures of misfitting, i.e. the lack of adaptation between subjects and model, were provided. An important information for the misfit diagnosis is given by the infit mean-squared statistic [11], which is an inlier-pattern-sensitive fit statistic. This is based on the chi-square statistic with each observation weighted by its statistical information (model variance) and is more sensitive to unexpected patterns of observations by persons on items (and vice-versa). More formally:

$$Infit_{ms} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{K} z_{ij}^2 \sigma_{ij}^2}{\sum_{i=1}^{n}\sum_{j=1}^{K} \sigma_{ij}^2} \quad (4)$$

where z_{ij} is the standardized residual variance across both persons and items and σ_{ij} is the weight represented by the individual variance.

3.2 The latent class model

Another approach we compared was a simultaneous modeling through PCM and clustering, which can be achieved assuming a Latent Class Model [8].

The basic idea underlying LC model is that the probability of obtaining a particular response pattern x, $P(X = x)$, is a weighted average of the K class-specific probabilities $P(X = x|\theta = t)$; that is,

$$P(X = x) = \sum_{\theta=1}^{K} P(\theta = t)P(X = x|\theta = t) \quad (5)$$

We assumed that subjects were divided into k clusters of (approximately) similar ability. Hence θ_i was considered as a discrete random effect which could take k possible values with a certain discrete probability distribution.

These values (latent cluster centroids) and their probability were estimated by an ad-hoc expectation maximization algorithm designed for this model.

Model choice can be performed by minimizing the Bayesian Information Criterion (BIC) [12]. Its general formulation is given by:

$$BIC = -2\ln(L) + k\ln(n) \quad (6)$$

where L is the maximized value of the likelihood function for the estimated model, k the number of free parameters to be estimated and n the sample size.

4. Results

Both the scales presented high reliability, with values of Cronbach's Alpha equal to 0.85 for PAQ and 0.94 for VascuQoL.

Table 1. Model measure and Infit statistic - PAQ.

Item	Model measure	Infit mnsq
Satisfaction for medical explanations	43.69	0.54
Limitation in doing hobbies or relaxing activities	43.45	1.87
Changing of symptoms	51.94	0.75
Satisfaction for the past attempts of therapy	49.56	0.57
Limitation in visiting friends or relatives	45.89	1.76
Satisfaction for the present therapy	49.1	0.42
Limitation in working or doing houseworks	48.51	1.35
Satisfaction for a possible symptoms stability	49.97	1.09
Limitation in running	56.09	1.9
Limitation in doing hard works or in working out	54.75	2.35
Depression for the pathology	51.46	1
Pain for cramps	49.56	0.98
Limitation in walking near home	50.17	0.54
Insomnia for pain or cramps	44.6	0.58
Limitation in enjoyments	50.78	1.05
Limitation in walking for 150 – 200 meters on plain	57.1	0.64
Frequency of pain or cramps	45.15	0.89
Limitation in walking for 50 – 100 meters on climb	55.73	0.5
Limitation in walking for 50 – 100 meters on plain	52.49	0.5

The correspondence analysis highlighted the presence of Guttman's effect, which can be observed by a paraboloid distribution of the categories on the factorial plane. This pointed out that data implied a unique latent variable which can be interpret as Quality of Life and satisfies one important PCM hypothesis.

Through the application of cluster analysis on factors, three categories of patients were identified: patients with good, moderate and bad Quality of Life. These groups of patients can also be clearly observed on the factorial plane, and give further evidence for the application of LC model.

The application of PCM identified some relevant changes in the questionnaires structure, among which there are:

- the elimination of some items, which makes the QoL measurement worse;
- the organization of the response structure into three categories, instead of five, six or seven.

The first result was made clear by the observation of infit statistic, the second one by the Category Probability Curves, which represent the probability of observing each ordered category as a function of the items measure of difficulty. For the interpretation of the infit statistic we used the "permissive" approach, which is more adequate for small sample, considering values between 0.6 and 1.4 as acceptable. In fact, when the infit statistic is distributed as mean squares, its expected value is 1, but guidelines for determining unacceptable departures from expectation consider an item or a person as misfit when the mean square infit value is larger than 1.4 for samples less than $n = 500$. Table 1 shows the estimated infit mean squared statistic and the model measure, representing an estimate of the items' difficulty parameter, for the items of PAQ questionnaire. We interpreted estimate of the model measure as the importance given by an individual to a particular item, expressed in the *logit* unit of measure.

Items which did not meet the infit statistic requirements are 'Limitation in hobbies', 'Limitation in visiting friends or relatives', 'Limitation in running' and 'Limitation in hard works', which presented values larger than 1.4; therefore these items were interpreted to make the measurement of Quality of Life worse and we suggested to eliminate them from the PAQ questionnaire. Explanations to this choice were that while the first two items were not discriminating, concerning easy activities (anyone could perform relaxing activities), the remaining ones suffered from the opposite effect.

Using the same criterion we gave an interpretation to the infit statistic computed for VascuQoL. Three items ('Cold feet', 'Swarming' and 'Fatigue') present values greater than 1.4, regarding widespread symptoms in PAD patients. Nonetheless we proposed not to remove them from the VascuQoL questionnaire, but to unify them, concerning important symptoms of PAD.

The categories structure according to the PCM was found inadequate for this particular sample. Figure 1 shows the Category Probability Curves for PAQ questionnaire, representing the probability of observing each

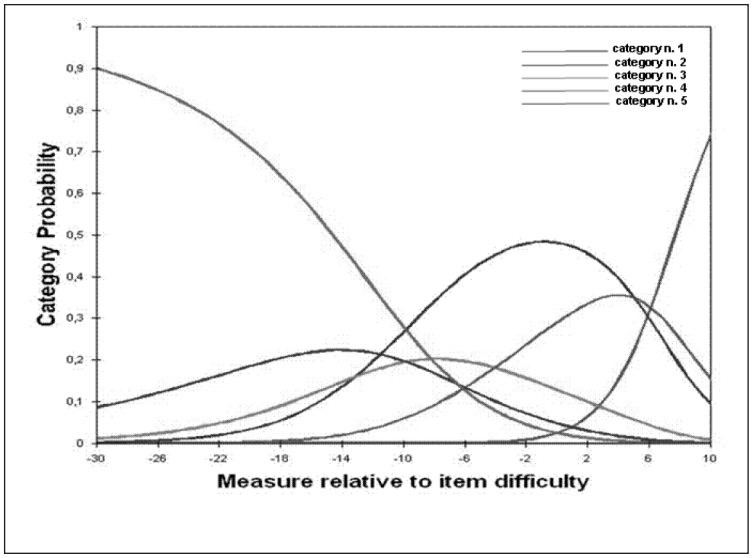

Figure 1. Category Probability Curves - PAQ.

ordered category as a function of items' difficulty.
The thresholds between the categories (represented by the point of intersection for any two curves) are very close together and their locations do not always increase according to the logical order principle required by the PCM. The distribution of categories is trimodal, suggesting that the correct number of categories according to this specific sample is three, instead of four or five. The same conclusions can be made for VascuQoL.
The LC model highlighted a three classes structure of the sample, selected by the BIC. The cluster are characterized as in Table 2. The overlap (measured with the percentages of individuals belonging to the same cluster) between the classifications obtained through the LC and the cluster analysis on factors is 47% for PAQ and 81% for VascuQoL. Anyway the values of the centroids allowed us to interpret the clusters in the same way for both the analysis, i.e. as different gradations of QoL.
Moreover it is interesting to note that while for VascuQoL questionnaire the best model resulted the simple PCM (suggesting that possibly there is no clear clustering), for PAQ the best model was the LCM with *k* = 3 clusters, which can have been easily characterized as bad, satisfactory and good quality of life.
Values of Bayesian Information Criterion are shown in Table 3.

Table 2. Cluster characterization (LC) - PAQ and VascuQoL.

	Cluster	n
PAQ	1	37
	2	8
	3	55
	1	19
V ascuQoL	2	62
	3	19

Table 3. Bayesian Information Criterion - PCM vs LC.

	PCM	LC2	LC3	LC4
V ASCUQOL	5603.60	5799.69	5602.15	5701.64
PAQ	7891.35	9365.81	8073.32	8093.12

5. Conclusions

Since peripheral arterial disease (PAD) is a common condition highly associated with cardiovascular morbidity, the evaluation of QoL through reliable instruments plays a central role in establishing whether the therapy improves the patients conditions.
Other than the classical measures of correlation for the evaluation of the reliability and the application of the correspondence analysis which highlighted the dimensions measured, our approach was based on the comparison of the PCM and the LC.
The results obtained assure that both PAQ and VascuQoL are characterized by good reliability and internal consistency (measured by Cronbach's Alpha), but require some improvements both in items and categories structure, according to the application of PCM and LC.

The Infit Statistics were observed in order to find out which items give the best measure of QoL. The values obtained suggest the elimination or the aggregation of some items.

The Category Probability Curves followed a trimodal distribution, suggesting that the most adequate number of response categories is three and not five, six or seven as the original formulation of the two questionnaires. Finally, the Bayesian Information Criterion (BIC) led to select the LCM with three classes as the best model for PAQ, and PCM as the best one for VascuQoL. In general, through the application of the two model, a clear clustered structure of the sample was appeared. Particularly the cluster analysis, the correspondence analysis, the category probability curves and the latent class model identified three clusters corresponding to different grade of QoL. This allowed us to argue that both the instruments can correctly measure the Quality of Life. We have suggested few adjustments for improving them.

References

1. Van Korlaar I. Vossen C. Rosendaal F. Cameron L. Bovill E. Kaptein A. Quality of life in venous disease. *Thrombosis and Haemostasis,* 90:27–35, 2003.
2. Liles D. R. Kallen M. Petersen L. Busch. Quality of life and peripheral arterial disease. *Journal of Surgical Research,* 136:294-301, 2006.
3. Dumville C. Lee A. Smith F. Fowkes F. Bovill E. Kaptein A. The health-related quality of life of people with peripheral arterial desease in the community: the edinburgh artery study. *British journal of general practice,* 54:826–831, 2004.
4. Nguyen L. Moneta G. Conte M. Bandyk D. Clowes A. Seely B. Prospective multicenter study of quality of life before and after lower extremity vein bypass in 1404 patients with critical limb ischemia. *Journal of Vascular Surgery,* 44:977–983, 2005.
5. Spertus J. Jones P. Poler S. Rocha-Singh K. The peripheral artery questionnaire: A new disease-specific health status measure for patients with peripheral arterial disease. *American Heart Journal,* 147:301–308, 2004.
6. Morgan M. Crayford T. Murrin B. Fraser S. Developing the vascular quality of life questionnaire: A new disease-specific quality of life measure for use in lower limb ischemia. *Journal of Vascular Surgery,* 33:679–690, 2001.
7. Met R. Reekers J. Koelemay M. Legemate D. Haan R. The amc linear disability score (alds): a cross-sectional study with a new generic instrument to measure disability applied to patients with peripheral arterial disease. *Health and Quality of Life Outcomes,* 7:88–97, 2009.
8. Goodman L.A. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika,* 61:215231, 1974.
9. Cronbach L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 16:297–333, 1951.
10. Geoff N. M. A rasch model for partial credit scoring. *Psychometrika,* 47:149–174, 1982.
11. Wright B. and Masters G. *Rating scale analysis.* Mesa Press, Chicago, 1982.
12. Schwartz G. E. Estimating the dimension of a model. *Annals of Statistics,* 6:461–464, 1978.