ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

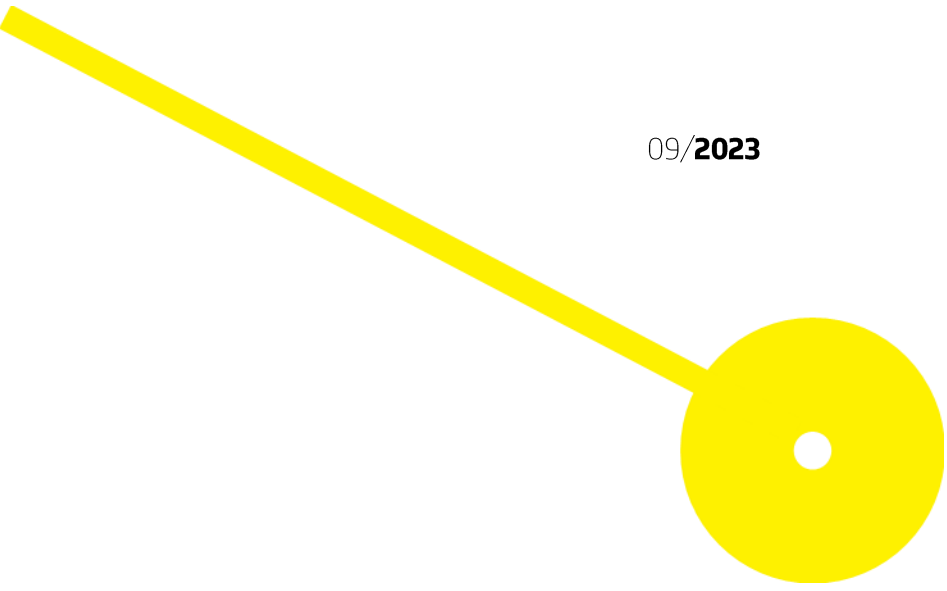
P.PORTO

M

MESTRADO
MESTRADO EM BIOESTATÍSTICA E BIOINFORMÁTICA APLICADAS À SAÚDE

Contributions for the Validation of the Portuguese version of the Vascular Quality of Life-6 Questionnaire in Peripheral Artery Disease Patients

Rafaela Monteiro Oliveira



09/**2023**

P.PORTO

ESCOLA SUPERIOR DE SAÚDE

Contributions for the Validation of the Portuguese version of the Vascular Quality of Life-6 Questionnaire in Peripheral Artery Disease Patients

Autor

Rafaela Monteiro Oliveira

Orientador(es)

Professora Doutora Ivone Silva / Centro Hospitalar Universitário de Santo António (CHUP)

Professora Doutora Susana Pedras / Centro Hospitalar Universitário de Santo António (CHUP)

Professor Doutor Rui Pimenta / Escola Superior de Saúde (ESS), Instituto Politécnico do Porto

Dissertação de Estágio apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em **Bioestatística e Bioinformática Aplicadas à Saúde** pela Escola Superior de Saúde do Instituto Politécnico do Porto.

Agradecimentos

Gostaria de agradecer aos meus orientadores, Professora Doutora Ivone Silva, Professora Doutora Susana Pedras e Professor Doutor Rui Pimenta, cujo apoio e orientação foram fundamentais ao longo deste percurso.

Ao Serviço de Angiologia e Cirurgia Vascular do Centro Hospitalar Universitário de Santo António, pela forma como me receberam e pela oportunidade de realizar este trabalho de investigação.

À equipa envolvida no projeto WalkingPad, do qual tenho a honra de fazer parte, expresso o meu profundo orgulho por contribuir para melhorar a qualidade de vida dos nossos pacientes.

E acima de tudo, a todos os pacientes que generosamente participaram no estudo, porque sem a sua colaboração, a realização deste trabalho não teria sido possível.

Ackowledgments

I would like to thank my supervisors, Professor Ivone Silva, Professor Susana Pedras and Professor Rui Pimenta, whose support and guidance were fundamental throughout this journey.

To the Angiology and Vascular Surgery Department of the Centro Hospitalar Universitário de Santo António, for the way they welcomed me and for the opportunity to carry out this research work.

To the team involved in the WalkingPad project, of which I am honored to be a part, I express my deep pride in contributing to improving the quality of life of our patients.

And above all, to all the patients who generously participated in the study, because without their cooperation, this work would not have been possible.

Resumo

A Doença Arterial Periférica (DAP) é uma doença aterosclerótica oclusiva que afeta >230 milhões de pessoas em todo o mundo. O sintoma mais comum é a claudicação intermitente (CI) que leva a uma diminuição da qualidade de vida (QV). Assim, este estudo teve como objetivo contribuir para a validação do questionário VascuQol-6 para a população portuguesa, de modo a obter uma forma rápida, sensível e de fácil utilização para avaliar a QV na DAP.

O VascuQoL-6 foi adaptado e traduzido para o português europeu. Foram incluídos 115 doentes com uma idade média de 65 anos e com DAP com CI estável há mais de 3 meses. Foram testadas a fiabilidade, a análise da validade de constructo através da validade convergente e discriminante, a validade do grupo conhecido (*known-group validity*) e a análise da responsividade.

A Variância Média Extraída para o construto latente foi de 0,40 e a Confiabilidade Composta foi de 0,79, indicando forte consistência interna. O VascuQoL-6 foi positivamente associado aos scores da Componente Física e da Componente Mental do SF-36 ($r = 0,64$, $p < 0,01$ e $r = 0,42$, $p < 0,01$, respetivamente). Por outro lado, não houve correlação significativa entre os scores do VascuQoL-6 e o PADKQ ou IPAQ. Foi encontrada uma diferença estatisticamente significativa entre os grupos de acordo com a gravidade da CI ($F(2,47) = 8,35$, $p < 0,001$). Um teste t para amostras emparelhadas mostrou diferenças estatisticamente significativas entre os scores do VascuQol-6 antes de um programa de caminhada ($M = 15,65$, $DP = 3,09$) e depois de um programa de caminhada ($M = 17,41$, $DP = 2,71$), $t(67) = 3,94$, $p = < 0,001$.

O VascuQoL-6 é um instrumento de 6 itens para avaliar a QV associada à DAP com boas propriedades psicométricas, validade convergente e discriminante com o SF-36, PADKQ e IPAQ. O instrumento demonstrou ter validade de grupo e capacidade de resposta conhecidas.

Palavras-chave: Doença Arterial Periférica; Qualidade de Vida; Validação; Fiabilidade

Abstract

Peripheral Arterial Disease (PAD) is an occlusive atherosclerotic disease that affects >230 million people worldwide. The most common symptom is intermittent claudication (IC) that leads to a lower quality of life (QoL). Thus, this study aimed to contribute to the validation of the VascuQol-6 questionnaire for the Portuguese population to obtain a quick, sensitive, and easy-to-use way to assess QoL in PAD.

The VascuQoL-6 was adapted and translated into European Portuguese. 115 patients were included with a mean age of 65 years and with PAD with IC stable for more than 3 months. Reliability, construct validity analysis through convergent and discriminant validity, known-group validity, and responsiveness analysis were tested.

The Average Variance Extracted for the latent construct was 0.40 and the Composite Reliability was 0.79, indicating strong internal consistency. VascuQoL-6 was positively associated with SF-36 Physical Component Summary and Mental Component Summary scores ($r = .64$, $p < .01$ and $r = .42$, $p < .01$, respectively). In turn, there was no significant correlation between VascuQoL-6 scores and the PADKQ or IPAQ. A statistically significant difference between groups according to IC severity ($F(2,47) = 8.35$, $p < .001$) was found. A paired samples t-test showed differences between VascuQol-6 scores before a walking program ($M = 15.65$, $SD = 3.09$), and after a walking program ($M = 17.41$, $SD = 2.71$), $t(67) = 3.94$, $p = < .001$.

The VascuQoL-6 is a 6-item instrument to assess the QoL associated with PAD with good psychometric properties, convergent and discriminant validity with SF-36, PADKQ and IPAQ. The instrument proved to have known group validity and responsiveness.

Keywords: Peripheral Arterial Disease; Quality of Life; Validation; Reliability

Abbreviations and Acronyms

ABI – Ankle-brachial Index

AVE – Average Variance Extracted

CfA – Confirmatory Factor Analysis

CR – Composite Reliability

EFA – Exploratory Factor Analysis

HRQoL – Health related quality of life

IC – Intermittent Claudication

ICC – Interclass Correlation Coefficients

IPAQ - International Physical Activity Questionnaire

MCS - Mental Component Score

PAD – Peripheral Artery Disease

PADKQ - Peripheral Arterial Disease Knowledge Questionnaire

PCS - Physical Component Score

PFWD – Pain Free Walking Distance

QoL – Quality of Life

SF-36 - Short-Form Health Survey

VascuQoL-6 – Vascular Quality of Life Questionnaire-6

Table of Contents

Tables

1. Introduction

Questionnaire validation is a pivotal step, for acquiring a questionnaire that is psychometrically sound, and is efficient and effective for use in research and clinical settings. This multifaceted task involves several intricate stages, each contributing to the overall quality and effectiveness of the research instrument(1). The present dissertation comprises two distinct parts, each addressing critical aspects of the questionnaire validation process.

The first part of this dissertation embarks on the process of questionnaire validation where a validation article intended for submission to the esteemed Acta Médica Portuguesa journal is presented, titled "Contributions for the Validation of the Portuguese version of the Vascular Quality of Life-6 Questionnaire in Peripheral Artery Disease Patients". This section showcases the steps and methodologies that were taken to validate a specific QoL questionnaire, giving insight into the limitations and successes of adapting and validating a questionnaire to a new cultural and linguistic contex in clinical settings.

In the second part of this dissertation, the focus shifted to the broader landscape of questionnaire validation methodologies. The principal methods that comprise the entire questionnaire validation process are explored. There are several different methods and strategies, designed to navigate the nuances of study settings, database characteristics, sample sizes, and the variables under evaluation(1). These section serves as a toolbox for researchers, allowing them to tailor their validation efforts to the unique demands of their studies.

2. Part I – Contributions for the Validation of the Portuguese version of the Vascular Quality of Life-6 Questionnaire in Peripheral Artery Disease Patients

2.1. Introduction

Peripheral Arterial Disease (PAD) is an occlusive atherosclerotic disease of the arteries of the extremities of the body that affects more than 230 million people worldwide (2,3). Intermittent Claudication (IC) is the most common symptom of PAD and is described as leg pain that occurs mainly while walking(4). These symptoms impair the ambulation and functional capacity of patients, leading to loss of mobility, disease deterioration, increased risk of other cardiovascular diseases, and lower quality of life (QoL)(5).

Disease severity and functional impairment in patients with IC are usually quantified by ankle-brachial index (ABI) at rest and after exercise and walking distance by standardized walking tests(6,7). However, the exclusive use of these parameters requires specialized equipment and professionals, is costly, time-consuming, and does not adequately reflect the patient's day-to-day walking ability and perception of their functional impairment(8). In addition, these parameters do not assess the impact of the disease concerning the social and emotional consequences of living with PAD. Therefore, it has been suggested that traditional measures should be complemented with patient-reported outcome measures, such as health-related QoL instruments(9).

Health-Related Quality of Life (HRQoL) questionnaires have been used as an essential outcome tool in the evaluation of clinical outcomes, aiming to assess the impact of the disease and treatment on the biopsychosocial scope of the patient. In addition, they are of great relevance in improving knowledge about different conditions and in analyzing the economic cost of an intervention(10).

In a systematic review by Poku et al.(11), different tools used for psychometric evaluation were analyzed and the Vascular Quality of Life Questionnaire (VascuQol) was recommended as the preferred questionnaire to measure HRQoL outcomes in patients with PAD(10,12).

The Vascular Disease-Specific Quality of Life Questionnaire (VascuQoL) is a 25-item specific measure of HRQoL for patients with PAD. A shorter version (VascuQoL-6) was developed to facilitate the use of the instrument, consisting of six items with different response scales, and is now widely used to assess QoL in patients with PAD(13,14). The validation study by Larsen et al.(14) demonstrated improved accuracy and discriminating power for measuring the QoL compared to the original 25-item questionnaire in a group of 21 patients with IC9.

This instrument was translated and validated in different countries, including in the Portuguese language for the Brazilian population (10,12,14,15). The Brazilian version of the VascuQoL-6 showed adequate valid and reliable indicators, allowing its use in patients with PAD and with IC symptoms(15).

Although the VascuQol-6 questionnaire has already been translated into Portuguese, it has only been validated in the Brazilian population, making its application in the Portuguese population difficult due to their sociocultural differences. Thus, this study aimed to perform a cross-cultural adaptation and validation of the VascuQol-6 questionnaire for the Portuguese population in order to obtain a fast, sensitive, and easily applicable way to assess QoL in Portuguese patients diagnosed with PAD.

2.2. Methods

A cross-sectional study of cultural adaptation and validation of the VascuQoL-6 questionnaire to the Portuguese population was conducted. All the patients provided written informed consent to participate in the study. This study was approved by the Department of Education, Training, and Research of Centro Hospitalar Universitário Santo António, on October 22, 2019 (reference no. 184/2020, of May 7, 2020, and reference 069-DEFI/068-CES, respectively).

The first phase of the study concerned the translation, cultural adaptation and verification of semantic equivalence of the questionnaire following the principles of good practice for the translation and cultural adaptation process of patient report outcomes(16,17).

The questionnaire was originally provided by the author in the target language (European Portuguese). As such, a direct translation of the questionnaire was not necessary. Instead, the focus was on cultural adaptation, which involved ensuring that the items of the questionnaire were clear, relevant, and appropriate for the study population. To this end, an expert committee composed by vascular surgeons and nurses were requested to verify and evaluate the content of the items, i.e., if the items were adequate and adapted to the population, to identify any potential cultural differences or ambiguities that could affect the interpretation of the questions(1). After the revision of each item, the expert committee concluded that no modifications were necessary, as the questionnaire items were deemed clear and relevant for the study population.

The second phase of the study concerned evaluation of reliability and validity of the translated version of the scale. This included the internal consistency analysis, construct validity analysis through convergent and discriminant validity, known-group validity, and responsiveness analysis.

Participants

The VascuQoL-6 questionnaire was applied to voluntary participants whose native language is European Portuguese. These participants were PAD patients followed in an Angiology and Vascular Surgery Department of a Central Hospital. Inclusion criteria were (a) Symptomatic PAD; (b) IC stable for more than 3 months; (c) Ankle–brachial index (ABI)<0.9 at rest; and (d) age between 50 and 80 years old. Were excluded from the study patients with cognitive impairment (evaluated through Mini Mental State Exam),

and asymptomatic PAD (to check the other inclusion and exclusion criteria, see xxx blinded for review purposes).

Data Collection

All data collection was performed during 2021 at Centro Hospitalar Universitário Santo António and was divided in two different evaluation sessions. The sessions were performed by a multidisciplinary team constituted by a vascular surgeon, a psychologist, a physical exercise technician and a clinical physiologist. The assessment was performed face-to-face in an interview format. In a first session, the ABI was measured using a handheld Doppler device (LifeDop 150 Doppler (8 MHz), USA), according to the AHA/ACC guidelines(18).

A treadmill test was conducted to provide an objective assessment of performance since it is an accepted method used in patients with IC to evaluate walking ability. Severity is generally described in terms of claudication distance, and it can be objectively measured by treadmill test with the assessment of pain-free walking distance (PFWD)(19,20). This way, the modified Gardner-Skinner Treadmill Protocol was used, where participants begin to walk on the treadmill at 1 km/h with a 0% grade. After 2 min, the speed is increased to 1.6 km/h, at 0% grade. Then, the speed is increased by 0.8 km/h every 2 min until reaching 3.2 km/h. After reaching 3.2 km/h, the speed is kept constant, and the grade is increased by 2% every 2 min(21). PFWD was recorded as the first sign of claudication pain.

To access reliability and validity of the scale, patients were asked to answer the VascuQoL-6, Short-Form Health Survey (SF-36), International Physical Activity Questionnaire for elderly-Short Form (IPAQ), Peripheral Arterial Disease Knowledge Questionnaire (PADKQ) questionnaires.

Participants that filled the inclusion criteria were prescribed with a physical exercise program, that included walking 3 times a week for a minimum of 30 minutes for the duration of 3 months, which is the first-line therapeutic measure recommended by clinical practice guidelines(18).

In a second session, after 3 months, the participants completed again to the VascuQoL-6 questionnaire allowing the assessment of the degree to which the VascuQoL-6 questionnaire is sensitive to change in response to an intervention (physical exercise program).

Questionnaires

Sociodemographic and clinical data questionnaire

The sociodemographic questionnaire included questions on gender, age, employment status, and education level, and clinical data questionnaire included information about risk factors and lifestyle habits.

VascuQoL-6 questionnaire

The VascuQol-6 questionnaire is composed of six items that reflect the five dimensions of the original VascuQoL and assess the QoL of patients with PAD. These six items relate to limitation in performing activities (activity), tiredness in the legs (symptom), walking ability (activity), concern about poor circulation in the legs (emotional aspect), ability to participate in social activities (social aspect), and discomfort from pain in the legs (pain)(12). Each item has a four-point response scale ranging from 1 (worst patient-perceived QoL) to 4 (best patient-perceived QoL). The responses to the items are summed to generate an overall score ranging from 6 to 24 points. Higher values indicate better health status(15).

Short-Form Health Survey (SF-36)

The SF-36 is a widely used generic HRQoL measure that encompasses multiple dimensions of physical and mental health. This instrument consists of 36 items with different response scales assessing eight health concepts: bodily pain, physical functioning, role limitations due to physical problems, mental health, vitality, social functioning, role limitations due to emotional problems and general health. Two summary components can be calculated: physical component score (PCS) and mental component score (MCS). The scores for each domain were transformed to a 0-100 scale, with higher scores indicating better HRQoL(9,22). In this sample, Cronbach alpha for PSC was 0.73 and for MCS was 0.84.

International Physical Activity Questionnaire for elderly-Short Form (IPAQ-SF-E)

The IPAQ version adapted for the elderly was used. This version consists of 4 self-reported moderate-to-vigorous physical activity and sedentary behavior (sitting) items. The items encompass the following behaviors, in the last 7 days: the time spent sitting, the days and time spent walking, the days and time spent in moderate-intensity activities, and the days and time spent in vigorous-intensity activities. Scores range from 0 to indefinite minutes of physical activity per week and higher results correspond to a greater amount of physical activity performed. Results are reported in categories (low, moderate, or high activity levels)(23).

Peripheral Arterial Disease Knowledge Questionnaire (PADKQ)

PADKQ is a 16-item questionnaire that assesses patients' level of knowledge about PAD risk factors, symptoms, treatment options, and self-management strategies. The total score ranges from 0 to 16 with higher scores on the PADKQ indicating greater knowledge about PAD(24). In this sample, Cronbach alpha was 0.77.

Statistical Analysis

Descriptive statistics were used for characterization of the sociodemographic variables of the study population and for analysis of the VascuQoL-6, SF-36, IPAQ, and PADKQ scores. The distribution analysis

of the quantitative variables was performed using the Shapiro-Wilk test, and all variables had a normal distribution. This way, parametric tests were used for the statistical analysis.

In terms of reliability, internal consistency reflects the extent to which the questionnaire items are inter correlated, or whether they are consistent in measurement of the same construct(1). Internal consistency of the VascuQol-6 items was assessed using the Cronbach's alpha, which is the most used internal consistency measure(25). Values higher than 0.5 indicate an acceptable level of reliability and values above 0.70 are a commonly used threshold for questionnaires intended for clinical use(26–29).

McDonald's Omega, Average Inter-item Correlation and Composite Reliability (CR) were also calculated since these are alternative measures to Cronbach's alpha. McDonald's Omega is based on a factor analytic approach and some authors defend that it has been proven to be more robust that Cronbach's alpha. Values of 0.70 or higher are considered acceptable(30). Average Inter-item Correlation involves calculating the correlation coefficients between each pair of items within the questionnaire and then computing the average of these correlations. Some authors defend that this measure is better for scales that have a small number of items (less than 10) which may invoke low Cronbach's alpha values. Optimal average inter-item correlation values range from 0.2 to 0.4 (31). Composite Reliability (CR) was also used for evaluating internal consistency since it is a less biased estimate of reliability than Cronbach's Alpha. A CR value of 0.70 or higher is generally considered acceptable(32).

Convergent and divergent validity are integral aspects of assessing the construct validity of a questionnaire, ensuring that it accurately measures what it is intended to measure(1,33). Pearson correlation coefficients were computed to evaluate (a) the relationship between the score obtained in the VascuQoL-6 questionnaire and the SF-36 questionnaire to assess convergent validity and (b) evaluate the relationship between the score obtained in the VascuQoL-6 questionnaire and the IPAQ and PADKQ to assess divergent validity. Regarding the sample size, 115 participants answered the VascuQoL-6 and the IPAQ, 111 patients completed the PADKQ and 49 completed the SF-36. Thus, analyzes with the SF-36 included only 49 patients.

Average Variance Extracted (AVE) was also calculated, since it is also a commonly used method to assess convergent validity. A threshold for AVE of 0.50 or higher is generally considered acceptable(32).

Known-groups validity is another way of assessing the validity of an instrument(34). For this, one-way analysis of variance (ANOVA) was conducted to compare the mean VascuQoL-6 scores among the PAD clinical presentation of participants. Patient treadmill PFWD was categorized based on the Fontain classification where patients with PFWD ≥200 meters were classified with Mild IC (stage IIa), 50-200 with Moderate IC (stage IIb) and <50 with Severe IC (stage III)(35).

Tukey's Post-hoc test, was performed to identify specific group differences. A p-value of <0.05 was considered statistically significant. The known-group validity of the VascuQoL-6 questionnaire was determined based on the ability of the questionnaire to differentiate between different clinical

presentation of PAD (Mild IC, Moderate IC, and Severe IC) as indicated by significant differences in mean scores across Treadmill PFWD categories.

Responsiveness is considered the longitudinal aspect of validity(36). To assess the responsiveness of VascuQoL-6 questionnaire, change scores were calculated by subtracting the baseline scores from the scores obtained after the 3-month physical exercise program. The statistical analysis included paired t-tests to determine the significance of changes in VascuQoL-6 scores following the intervention consisting in a physical exercise prescription. A p-value of <0.05 was considered statistically significant. Effect sizes (Cohen's criteria) were also calculated to quantify the magnitude of change. These effect sizes were evaluated based on established benchmarks: a small effect size was indicated by d = 0.2, medium by d = 0.5, and a large by d ≥ 0.8(37). The responsiveness of the VascuQoL-6 questionnaire to the physical exercise program was evaluated based on the magnitude and significance of changes in scores over time.

2.3. Results

From the 115 participants that answered to the VascuQoL-6, 14.8% were female and mean age was 65 years. Most of the patients has hypertension (86.1%), dyslipidemia (87.8%), and were current or previous smokers (91.3%). Detailed demographic and risk factor characteristics can be found in Table 1.

Table 1. *Demographics and risk factor characteristic in patient population (N=115)*

Variables		**Mean (SD) or n (%)**
Sociodemographic		
Sex	Male	98 (85.2)
	Female	17 (14.8)
Age (years)		64.67 (7.23)
Education level (years)		6.41 (3.77)
Risk Factors		
Hypertension	No	16 (13.9)
	Yes	99 (86.1)
High Cholesterol	No	14 (12.2)
	Yes	101 (87.8)
Obesity (BMI>30)	No	85 (73.9)
	Yes	30 (26.1)
Diabetes *mellitus* type 2	No	65 (56.5)
	Yes	50 (43.5)
Smoking History (active or former)	No	10 (8.7)
	Yes	105 (91.3)
Ankle-brachial index (ABI)		
Right ABI		0.71 (0.19)
Left ABI		0.71 (0.18)

Table 2 provides a comprehensive characterization of the evaluation instruments employed in this study.

Table 2. Characterization of evaluation instruments (N=115)

Variables		Mean (SD) or n (%)
Treadmill Test[a]		
Pain Free Walking Distance (PFWD) (meters)		125.31 (118.31)
Fontain Classification based on PFWD	Mild Claudication – stage IIa (≥200 meters)	26 (24.3)
	Moderate Claudication – stage IIb (50–200 meters)	45 (42.1)
	Severe Claudication– stage III (<50 meters)	36 (33.6)
VascuQoL-6		
Item 1: Activity		3.17 (0.89)
Item 2: Symptoms		2.25 (1.22)
Item 3: Activity (walking)		2.59 (0.67)
Item 4: Emotional		1.96 (1.05)
Item 5: Social activities		3.50 (0.82)
Item 6: Pain		1.81 (0.70)
Total Score		15.28 (3.28)
SF-36[b]		
Physical functioning		18.69 (5.48)
Physical role limitations		11.55 (5.80)
Emotional role limitations		12.06 (3.84)
Mental health		17.57 (4.92)
Social functioning		7.98 (2.37)
Energy/vitality		14.29 (3.74)
General health perceptions		14.70 (3.50)
Bodily pain		5.75 (2.42)
Physical Component Score (PCS)		50.70 (13.43)
Mental Component Score (MCS)		51.90 (12.56)
IPAQ		
IPAQ categories	Insufficiently active	80 (69.6)
	Moderately active	29 (25.2)
	Vigorously active	6 (5.2)
Disease knowledge (PADKQ)[c]		11.07 (3.26)

[a] N = 107
[b] N = 49
[c] N = 111

To gain a deeper understanding of the QoL within this cohort, an analysis examining potential differences related to gender, age, and education in relation to the VascuQoL-6 total scores was conducted. Our analysis did not reveal significant differences in VascuQoL-6 scores among different demographic subgroups. The results of this analysis are presented in Table 3.

Table 3. Quality of Life Differences Among Demographic Subgroups

Variables groups	N	Mean (SD)	p-value
Male	98	15.50 (3.325)	0.082
Female	17	14.00 (2.784)	
Age ≤ 65	64	15.11 (3.019)	0.539
Age < 65	51	15.49 (3.608)	
Education Level ≤ 4 years	63	15.09 (3.627)	0.513
Education Level > 4 years	52	15.50 (2.832)	

Internal Consistency

The value obtained for the Cronbach's alpha was 0.644, indicating an acceptable internal consistency(27). McDonald's Omega was also measured, and the value obtained was 0.637.

Analyzing the internal consistency coefficients of Cronbach's alpha and McDonald's Omega presented above, it is possible to see that this scale is below the value of 0.70, a particularly commonly used threshold for questionnaires intended for clinical use(26,28,29). In this sense, the average inter-item correlation was calculated and revealed an average inter-item correlation of 0.266, suggesting an acceptable level of inter-item correlation, reflecting a reasonable degree of homogeneity among the items(31,38).

Composite Reliability (CR) was also calculated as another measure of internal consistency. The CR value obtained was 0.79, indicating strong internal consistency and reinforcing the questionnaire's reliability in measuring the intended construct(32).

Construct Validity

Convergent Validity

The AVE value obtained was 0.40, indicating that the collective variance explained by the items was slightly below the commonly recommended threshold of 0.50 for adequate convergent validity(32).

Also, to assess the convergent validity, the correlation between the VascuQoL-6 questionnaire and the SF-36 questionnaire was examined. The sample comprised 49 participants with PAD who completed both questionnaires. Pearson correlation coefficients were computed between the two measures. The Physical Component Summary (PCS) and Mental Component Summary (MCS) scores of the SF-36 correlated with the VascuQoL-6 scores ($r = .64$, $p < .01$ and $r = .42$, $p < .01$, respectively). Table 4 shows the correlation between the VascuQoL-6 score and SF-36 domains.

Discriminant Validity

To assess discriminant validity, the correlation between VascuQoL-6 scores and two other measures (PADKQ and the IPAQ questionnaire) was examined. Results indicated that there was no significant correlation between VascuQoL-6 scores and scores on either the PADKQ questionnaire or the IPAQ questionnaire. Table 4 shows the correlation between VascuQoL-6 score and PADKQ and IPAQ.

***Table 4**. Correlation coefficients (r) for physical and mental components of SF-36, IPAQ, and PADKQ scores and individual items of VascuQoL-6 at baseline.*

		Item 1: Activity	Item 2: Symptoms	Item 3: Walking Activity	Item 4: Emotional	Item 5: Social activities	Item 6: Pain	**Total score**
SF-36	Physical component score (PCS)	,380*	,274	,441*	,366*	,303*	,518**	**,635****
	Mental Component score (MCS)	,255	,180	,272	,288*	,106	,397*	**,419***
IPAQ		,067	,306	,083	-,144	-,109	-,073	**,065**
PADKQ		-,172	,199	,062	-,369	,014	,022	**-,078**

*p<0.05, **p<0.01

Known-Group Validity

To access Known-group validity, a one-way ANOVA was conducted to compare the mean VascuQoL-6 scores among the PAD clinical presentation (Mild IC, Moderate IC, and Severe IC). There was a statistically significant difference between groups ($F(2,47) = 8.35$, $p < .001$).

A Tukey post hoc test showed that the Mild IC group was able to walk without pain statistically significantly further than the Moderate IC group ($p = .005$) and Severe IC group ($p < .001$). There was no statistically significant difference between the Moderate IC and Severe IC groups ($p = .691$).

Responsiveness

A paired samples t-test was performed to evaluate whether there was a difference between VascuQol-6 scores before the walking program and after the walking program. The results indicated that the VascuQoL-6 summary score after the physical exercise program ($M = 17.41$, $SD = 2.71$) was significantly higher than the VascuQoL-6 summary score before the physical exercise program ($M = 15.65$, $SD = 3.09$), $t(67) = 3.94$, $p = < .001$. There was also a statistically significant improvement in the symptoms, walking ability and pain items (item 2, 3 and 6).

Size effect of SF-36 domains and component summary scores and all items and summary score of VQ-6 are shown in Table 5. According to Cohen's criteria, there was a moderate to large effect for Symptoms, Activity and Pain items, and a moderate effect size for VQ-6 summary score (d=0.5)(37).

***Table 5**. Responsiveness to change. Size Effect of SF-36, domains and component summary scores, and VQ6, all items and summary score.*

		Effect Score
VascuQoL-6	Activity	0,1
	Symptoms	0,5
	Walking Activity	1,6
	Emotional aspects	0,1
	Social aspects	0,2
	Pain	0,3
	VascuQoL-6 summary score	0,5
SF-36	Physical functioning	1
	Physical role limitations	0,5
	Emotional role limitations	0,3
	Mental health	0,8
	Social functioning	0,3
	Energy/vitality	0,7
	General health perceptions	0,2
	Bodily pain	1,1
	Physical Component summary score (PCS)	0,9
	Mental Component summary score (MCS)	0,7

2.4. Discussion

In this study of the psychometric properties for the HRQoL questionnaire VascuQoL-6 revealed an acceptable internal consistency, suggesting that the items within the questionnaire are sufficiently interrelated in measuring the construct of interest(27). The observed internal consistency, while deemed satisfactory, also revealed that other validation studies conducted on the VascuQoL-6 reported higher alpha values, indicating stronger inter-item reliability in those investigations(10,12,14,15). One plausible explanation for the differences might be the variations in the study populations across different validation studies. In our study, we deliberately excluded patients with critical ischemia, which represents the most severe form of PAD. By excluding this subset of patients, who often face unique challenges and significant impact on their QoL, we may have inadvertently influenced the internal consistency results of the VascuQoL-6.

In this sense, the average inter-item correlation was calculated and revealed a moderate level of inter-item correlation, reflecting a reasonable degree of homogeneity among the items. The average inter-item correlation provides valuable insights into the internal consistency of the VascuQoL-6 questionnaire, indicating that the items collectively contribute to a coherent measurement of peripheral artery disease-specific QoL(31,38).

The VascuQoL-6 questionnaire's internal consistency and convergent validity was examined through the evaluation CR and AVE, respectively. The CR value of 0.79 demonstrated strong internal consistency, reinforcing the questionnaire's reliability in measuring the intended construct. However, the AVE value for the latent construct was found to be 0.40, indicating that the collective variance explained by the items was slightly below the commonly recommended threshold of 0.50 for adequate convergent validity (32).

A good correlation between the physical domains and the PCS of the SF-36 and the VascuQoL-6 score was found, providing further evidence of convergent validity, although the correlation between the VascuQoL-6 and the MCS was low (0.42), which may be due to the fact that VascuQoL-6 includes only one item that assesses emotional aspects. These results suggest that the VascuQoL-6 questionnaire is measuring a construct similar to the SF-36 and that it is a valid measure of QoL in individuals with vascular disease. Similar results were demonstrated in previous validation studies of the VascuQoL-6 instrument, where correlations between the dimensions of SF-36 and the items in the VascuQol-6 were somewhat stronger for the items representing physical components(10,12,14). However, the AVE value for the latent construct was found to be 0.40, indicating that the collective variance explained by the items was slightly below the commonly recommended threshold of 0.50 for adequate convergent validity(32).

No significant correlation between VascuQoL-6 scores and scores on either the PADKQ questionnaire or the IPAQ questionnaire was found. These findings suggest that the VascuQoL-6 is measuring a distinct construct from disease knowledge and physical activity, providing evidence for discriminant validity of the questionnaire.

In the present study, we aimed to assess the known-group validity of the VascuQoL-6 questionnaire by comparing its scores among three groups of patients with varying degrees of IC severity: mild, moderate, and severe. Our findings revealed that the VascuQoL-6 questionnaire demonstrated significant discriminative ability in distinguishing between patients with mild and moderate IC, as well as between those with mild and severe IC. These results suggest that the questionnaire is effective in capturing meaningful differences in vascular-related QoL between these two pairs of groups, indicating its sensitivity to varying degrees of disease impact on patients' daily lives.

When comparing patients with moderate and severe claudication, the known-group validity analysis did not yield significant differences in VascuQoL-6 scores between these two groups. This unexpected finding warrants further exploration and consideration. One possible explanation could be the overlapping symptomatology and functional limitations experienced by patients in the moderate and severe claudication groups. It is plausible that patients in both groups may experience comparable levels of impairment, leading to similar VascuQoL-6 scores. Additionally, the VascuQoL-6 questionnaire may have limitations in distinguishing the subtle differences in QoL experienced by patients with moderate versus severe claudication, particularly if the impact of the disease becomes more profound in both groups.

Despite the non-significant result in the comparison between moderate and severe IC groups, the overall findings provide valuable insights into the known-group validity of the VascuQoL-6 questionnaire. The significant differences observed between mild and moderate IC, as well as between mild and severe IC, indicate that the questionnaire is capable of capturing clinically meaningful distinctions in QoL in relation to disease severity. As such, the VascuQoL-6 remains a valuable tool for assessing the impact of PAD on patients' QoL, particularly in distinguishing between patients with mild disease and those with more pronounced impairment.

In the present study, we aimed to evaluate the responsiveness of the VascuQoL-6 questionnaire to an intervention involving the prescription of physical exercise for patients with PAD. Our results demonstrated promising evidence of the questionnaire's sensitivity to changes in vascular-related QoL following the physical exercise program.

Specifically, the overall responsiveness analysis revealed a moderate effect size for the VascuQoL-6 summary score, indicating a meaningful and noticeable improvement in patients' overall QoL after engaging in the prescribed physical exercise program. This finding suggests that the VascuQoL-6 can detect clinically important changes in vascular-related QoL, highlighting its relevance as a valuable outcome measure in the context of interventions targeting PAD. Larsen et al., 2018 and Soria-Juan et al., 2021 also had similar findings where excellent responsiveness to change was demonstrated(10,14).

2.5. Conclusion

In conclusion, this study presents the Portuguese version of the VascuQoL-6 questionnaire and contributes to its validation as an instrument to assess the QoL of Portuguese patients with PAD. This tool can be especially valuable in follow-up evaluations, to measure the result of the physical exercise prescription, more invasive surgical interventions, as well as to compare the results with the international literature.

2.6. Study Limitations

The fact that the sample was collected in only one hospital in the north of the country does not allow us to generalize the results to other geographic areas. Moreover, the study cohort consisted primarily of male participants, potentially introducing a gender-related bias that may impact the generalizability of results. Another limitation is the variation in sample size, particularly the fact that only 49 patients completed the SF-36 questionnaire.

One significant limitation to acknowledge is the absence of a test-retest analysis. Ideally, such an analysis would have been valuable to assess the stability of responses over time. However, due to the three-month interval between assessments and the concurrent implementation of a physical exercise intervention, conducting a reliable test-retest analysis was not feasible within the constraints of the study. This extended interval and intervention may have introduced variability in participants' responses, potentially affecting the reliability of the instrument.

Additionally, the exclusion of patients with critical ischemia from the sample limits the generalizability of the results to this specific subgroup of individuals with PAD.

To address these limitations in future research, it is recommended that studies aim to recruit a more diverse and larger sample to enhance the external validity of the findings. While conducting a traditional test-retest analysis may not been feasible in our study contexts, researchers can explore this method to assess the stability of responses over time. Additionally, investigating the instrument's performance in a broader range of PAD severity levels, including patients with critical ischemia, will provide a more comprehensive understanding of its utility in clinical practice.

3. Part II – Methods for Questionnaire Validation in Clinical Research

3.1. Introduction

Questionnaires stand as a ubiquitous and indispensable instrument within the realm of research, bearing primary utility in the collection of data across various academic disciplines, encompassing the social sciences, medicine, psychology, economics, and beyond. This ubiquity is rooted in their reliability and capacity to systematically extract information in a standardized manner (39).

In the annals of scientific research, questionnaires are one of the dominant tools for data acquisition. Questionnaires are used to collect the relevant information from the respondents in order to measure each of the variables or items constituting its domain of interest, and also for addressing the objectives of a research study (40).

Questionnaires also facilitate the measurement of a vast range of parameters, from empirical facts to nuanced subjective experiences. That's why within the realm of healthcare and medicine, domains where precision and accuracy bear utmost significance, questionnaires emerge as invaluable tools for data collection. They fill in the gap between the subjective experiences of patients and the objective clinical data that underpins scientific inquiry (1).

However, it is extremely important for a researcher to know the importance of a proper questionnaire and whether it measures what it is intended to measure. In this sense, the intricacies of questionnaire design and the challenge of validation come to the fore (41).

This section embarks on an exploration of the process of questionnaire validation, delving into the multifarious methods and strategies available to ensure these instruments dutifully fulfill their role in the pursuit of empirical knowledge. This report aims to provide a straightforward guide for the development or translation of questionnaires and the validation process for use in clinical research.

3.2. Questionnaire development and questionnaire translation

This section discusses the steps that investigators should follow in developing or translating a questionnaire.

Firstly, it is crucial to identify the construct that is to be assessed with the questionnaire, as the domain of interest will determine what the questionnaire will measure (1). Once this construct has been delineated, a literature review should be undertaken. This review aims to find previously validated questionnaires applicable to comparable research settings, capturing variables that align with the study's hypotheses. In cases where existing questionnaires are unavailable or are deemed unsuitable, it is appropriate to construct a new questionnaire. Alternatively, if a suitable questionnaire exists but is in a different

language, the challenge lies in the translation and subsequent validation of the questionnaire in the target language (1,41).

3.2.1. Questionnaire development

Before writing the questionnaire items, several considerations must be considered.

Dimensionality of the construct

When embarking on the development of a questionnaire, the process entails the generation of items or questions designed to obtain participant responses, which are subsequently transformed into numerical data for rigorous statistical analysis. These items must reliably operationalize the core concepts outlined in the specific research questions and must be pertinent and well-received by the target demographic (33). Many constructs encountered in research are multidimensional, comprising various interrelated components. To fully assess the construct, it may be advisable to create subscales that assess each of these distinct components separately.

Subsequently, the weight of the questions also demands consideration. In cases where all dimensions of a construct bear equal importance, the same weight can be assigned to all questions. This can be accomplished by summing the responses or computing the average of all items. Conversely, when certain dimensions hold greater importance than others, it should be considered examining the results from each dimension separately (1).

Questionnaire mode of administration

In regards of questionnaire administration, two primary modes are commonly employed: a) self-administered and b) interviewer-administered questionnaires (1).

Self-administered questionnaires require only the distribution of the questionnaire itself. It offers economic advantages, as it does not require trained personnel, enabling the survey to reach a large and geographically diverse sample. This mode is particularly adept at capturing data on sensitive topics. This mode is less prone to information bias and interviewer effects but may face a higher risk of encountering missing responses. Common methods for distributing self-administered questionnaires include mail or electronic distribution.

On the other hand, the interviewer-administered approach involves direct interaction between interviewers and participants. Although it incurs higher costs, this method offers a valuable opportunity for real-time engagement. In recent times, telephone-based questionnaire administration has gained popularity as a cost-effective alternative and facilitates the inclusion of a wide-ranging and geographically diverse participant pool (41).

This decision depends, in part, on what the questionnaire intends to measure. For instance, when assessing constructs like catastrophic thinking related to pain, respondents may be more inclined to provide honest responses when completing the questionnaire independently, rather than when questioned directly by research or clinical staff. Another example is when, in scenarios where the questionnaire aims to measure patients' post-surgery mobility, respondents may be inclined to overstate their level of mobility as a sign of recovery (1).

Item generation and format

Question types play a pivotal role in shaping the nature and quality of data collected in a questionnaire. Two fundamental categories are open-ended and close-ended questions, each offering distinct advantages and challenges.

Open-ended questions allow respondents to provide detailed elaborations in their responses. These questions are best suited for situations where researchers seek in-depth insights into specific domains. However, open-ended questions often pose challenges in coding and scoring, making it more difficult to summarize individual responses.

Close-ended questions, provide respondents with a limited set of response options. Compared to open-ended questions, these items easier to administer and analyze. Nevertheless, the constrained response options may limit respondents' ability to clarify their answers, potentially influencing their responses based on the choices provided (1).

For close-ended items, the selection of an appropriate scale and response style becomes pivotal in questionnaire development. Various response styles serve distinct purposes, bearing different types and levels of data that subsequently influence the analysis options. Some response styles include (33):

- Frequency: Useful for establishing the frequency of a particular behavior or event. An example of a questionnaire that uses this type of response is the Intensive Care Experience (ICE) questionnaire or the Medida De Adesão Aos Tratamentos (MAT) questionnaire(42,43).
- Thurstone: Employs empirical data from judges to position attitudes or behaviors along a continuum with equal weighting and spacing. This type of response style is not commonly used in clinical assessment settings. An example of a questionnaire that uses this response style is the Nottingham Health Profile (44).
- Guttman: Utilizes hierarchical ranking to ensure that individuals agreeing with a particular item also agree with lower-ranked items. An example of a questionnaire that uses this response style is the Katz Index of Activities of Daily Living (45).
- Likert-type: The most common choice, employing fixed-choice response formats to measure attitudes or opinions. These ordinal scales gauge levels of agreement or disagreement, assuming

a linear continuum from strong agreement to strong disagreement, often offering respondents a choice of five, seven or nine pre-coded responses with a neutral point.

- Multiple choice/yes-no responses: Effective for evaluating the outcomes of educational programs, including patient education.

To facilitate subsequent statistical analyses, it is essential to scale items in a manner that generates sufficient variance among the target population. Questions should embody simplicity, clarity, and brevity, using precise language and minimal verbiage to inquire only about pertinent information. Lengthy or convoluted questionnaires can confuse both respondents and interviewers, potentially compromising data accuracy and provoking missing responses (41).

Ensuring face and content validity involves drawing from multiple sources, including expert consultation, input from proposed respondents, and a comprehensive review of relevant literature (33).

A number of guidelines have been suggested for writing items (46). Simplicity, conciseness, and use of familiar language for the target audience should be taken into consideration when constructing the items. Items should address single issues and avoid 'double-barreled' inquiries that merge multiple topics (e.g., "My disease affects my mental health and daily activities."). Leading questions should be avoided to prevent biased responses and questions expected to elicit uniform responses from all participants should not be used (e.g. "I would like that my health would improve"), as limited variance offers little insight into the construct being assessed.

The inclusion of reverse-scored items remains a topic of debate. Researchers opting for reverse-scored items should take extra precautions to ensure respondents interpret them as intended, maintaining their psychometric integrity alongside regularly coded items (1).

Review and revise initial pool of items

Upon the completion of the initial set of questionnaire items, it is imperative to subject them to a rigorous evaluation process by qualified experts. This review serves a multi-faceted purpose, encompassing the verification of item accuracy, the identification and rectification of any item construction issues, and the confirmation of grammatical correctness. Additionally, the reviewers must exercise due diligence in scrutinizing the items to prevent the inclusion of content that might be interpreted as offensive or biased by specific subgroups of respondents (1).

Preliminary testing

Prior to initiating a pilot test of the questionnaire with the designated respondents, it is prudent to subject the questionnaire items to a preliminary examination using a small sample, typically ranging from 30 to 50 respondents. This preliminary step serves several essential purposes. Firstly, it allows the questionnaire developer to assess whether any items elicit confusion among respondents and whether

participants offer valuable suggestions for potential item enhancements. Additionally, this phase provides an initial glimpse into the distribution of responses for each item, aiding in the determination of whether there exists adequate response variation to justify proceeding with a comprehensive large-scale pilot test. The assessment of feasibility, coupled with the identification of floor effects (where the majority of respondents score near the lower end) or ceiling effects (where the majority score near the upper end), assumes paramount importance during this stage. These factors play a pivotal role in deciding which items warrant inclusion and which should be reconsidered or discarded.
The questionnaire items should be revised upon reviewing the results of the preliminary pilot testing. This process may be repeated a few times before finalizing the final draft of the questionnaire (1).

3.2.2. Questionnaire translation

The translation of a questionnaire entails more than the translation of words. Questionnaires are designed to assess latent constructs that are not directly observable, making the quality of questionnaire items critical in ensuring that the intended construct is accurately measured and comprehended by the target audience. Both are crucial to generate valid results from which meaningful conclusions can be drawn. Errors in translation can introduce inaccuracies, potentially leading to misleading findings and, in the worst case wrong conclusions and decisions in both research and practical applications. Therefore, effective translation necessitates a careful consideration of linguistic attributes, such as word order and sentence structure, as well as cultural context, encompassing the nuanced connotations of words and idioms. Additionally, a deep understanding of the population for whom the questionnaire is intended is essential to ensure the questionnaire resonates effectively with its audience (47).
In the realm of questionnaire translation, two primary methods are employed, often in conjunction, to ensure linguistic and conceptual accuracy: forward translation and backward translation (1).

Forward Translation

The process of forward translation involves the initial translation of a questionnaire from its original language to the target language. This translation should be done by at least two independent bilingual translators, both proficient in the target language. ne of the translators should have a strong understanding of the concepts that the questionnaire intends to measure, which helps aligning the translation more closely with the original instrument. Simultaneously, a "naïve" translator, unaware of the questionnaire's objectives, performs the second translation. This approach helps to identify any differences or discrepancies between the original questionnaire and its translations. Any disparities between the translations can be discussed and resolved among the original translators, or with the inclusion of an unbiased, bilingual translator who was not involved in the initial translations.

Backward Translation

Once a questionnaire has been translated, the questionnaire is independently back-translated, which means translating it back from the target language into the original language. This step is crucial for verifying the accuracy of the translation and for uncovering any misunderstandings or ambiguities that may have arisen during the forward translation. It's important to use at least two independent translators for this step, who are preferably fluent in the original language. It is advisable that these back-translators remain unaware of the specific concepts the questionnaire aims to measure to prevent bias (1).

The International Test Commission's guidelines on translating and adapting tests endorse both forward and backward translation processes. These guidelines acknowledge the limitations of using forward and backward translation in isolation and recommend a combination of multiple translation approaches to overcome these limitations (47).

TRAPD approach

The TRAPD approach, represents a structured and comprehensive methodology for questionnaire translation and adaptation. The acronym TRAPD delineates the key steps of this translation process (47):

Translation: The first step is to accurately forward translate the questionnaire from the source language to the target language using bilingual translators who work independently to ensure accuracy.

Review: After translating the questionnaire, it undergoes a thorough review by an expert committee to evaluate its linguistic and conceptual fidelity. Members of the committee should include experts who are familiar with the construct of interest, a methodologist, both the forward and backward translators, and if possible, developers of the original questionnaires (1,47).

Adjudication: During the review, if any discrepancies or ambiguities are found, an adjudication step is introduced to resolve them. This phase helps to reconcile any differences between the original and translated versions by involving experts or a committee who typically oversee this process.

Pretest: Before the translated questionnaire is finalized, a pretest is carried out. This involves conducting the translated questionnaire on a small sample of individuals (about 30–50) who represent the target population. The goal of the pretest is to identify any potential issues with question comprehension, cultural relevance, or wording that may need further refinement. The respondents are asked to explain their understanding of each item in the translated questionnaire. This helps the investigator ensure the meaning remains the same as the original. This process may be repeated to finalize the translated version (1,47).

Documentation: The final step involves comprehensive documentation of the entire translation and adaptation process. This documentation includes detailed records of each step taken, any modifications made, and the rationale behind these decisions. Such documentation is crucial for ensuring transparency, replicability, and the traceability of the questionnaire's evolution.

3.3. Questionnaire Validation

Once the new or translated questionnaire items have undergone preliminary pilot testing and subsequent refinements, the next phase involves conducting a pilot test among the intended respondents for initial validation. In this pilot test, the final version of the questionnaire is administered to a substantial and representative sample of respondents who align with the questionnaire's intended demographic. It is worth noting that when pilot tests are conducted with small sample sizes, the relatively large sampling errors may reduce the statistical power needed to validate the questionnaire (1).

Validity is a crucial concept in questionnaire development. It refers to the extent to which the data collected accurately covers the intended area of investigation. In simple terms, validity means "measuring what is intended to be measured" (40).

Essentially there are three types of validity i) content validity, ii) criterion- related validity, and iii) construct validity. Also, reliability analysis is a crucial step for questionnaire validation.

A valid questionnaire helps the collection of better quality data which increases data credibility. A valid questionnaire must encompass the following characteristics: (a) simplicity and viability (b) reliability and precision in the words (c) adequate for the problem intended to measure (d) reflect underlying theory or concept to be measured and (e) capable of measuring change(41) .

In the following sections, the main types and methods for validation of a questionnaire are discussed.

3.3.1. Reliability

Reliability analysis is a crucial component of questionnaire validation. The reliability of a questionnaire can be considered as the consistency of the survey results(1) . Reliability refers to the repeatability, stability or internal consistency of a questionnaire (33). Ensuring that a questionnaire consistently measures the same underlying construct over time and across different situations or observers is critical for achieving a high degree of reliability. Confirming reliability is essential because it underpins the credibility and reproducibility of research findings, allowing researchers to draw meaningful conclusions and make informed decisions based on the data collected (25).

The reliability of a questionnaire can be evaluated using its internal consistency, test-retest reliability, and inter-rater reliability (1).

Internal Consistency

Internal consistency reflects the extent to which the questionnaire items are inter-correlated, or whether they are consistent in measurement of the same construct (1).

Cronbach Alpha coefficient is the most commonly used internal consistency measure. It is viewed as the most appropriate measure of reliability when making use of Likert scales (25).

Given a questionnaire x, with k number of items, alpha (α) can be computed as:

$$\alpha = \frac{\kappa}{\kappa - 1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2}\right)$$

Where, σ_i^2 is the variance of item i, and σ_x^2 is the total variance of the questionnaire.

Cronbach's alpha normally ranges from 0 to 1, with higher values indicating that items are more strongly interrelated with one another. Hinton et al. (2004) have suggested four cut-off points for reliability, which includes excellent reliability (0.90 and above), high reliability (0.70–0.90), moderate reliability (0.50–0.70) and low reliability (0.50 and below) (25). However, in practice, Cronbach's alpha of at least 0.70 has been suggested to indicate adequate internal consistency, especially for questionnaires intended for clinical use (26–29).

A low value Cronbach's alpha often suggests inadequate inter-relatedness among questionnaire items. In these cases, it is recommended to discard or revise items with low correlations with the questionnaire total score. If deleting the item reduces the Cronbach's alpha, this signifies the item should be retained. If deleting the item increases the Cronbach's alpha, this signifies the item should be removed. Also, Alpha is dependent on the length of the questionnaire, so it will increase with the number of items (39,48).

Alpha above 0.90 may suggest that some items are redundant as they are testing the same question but in a different guise. A maximum alpha value of 0.90 has been recommended (48).

Negative Cronbach's alpha values may occur when some items are negatively correlated with other items in the questionnaire. When reverse-scored items have not been appropriately reverse scored, it can be easily remedied by correctly scoring the items. However, if a negative Cronbach's alpha is still obtained even after correcting the scoring, it may signal fundamental issues within the questionnaire's original design (1).

There are alternative measures to Cronbach's alpha, such as **McDonald's omega**. Unlike alpha, which relies mainly on the correlation between the questions, McDonald's omega is based on a factor analytic approach. Some authors defend that Omega has proven to be more robust than alfa, although the difference between alpha and omega will often be small. Omega will have a value between 0 and 1. Internal consistency is also usually considered acceptable if the estimate is 0.70 or higher (30).

Montagni et all 2022, is an example of a study that accessed both measures, Cronbach's alpha and MacDonald's omega, for the reliability analysis of the Mental Health Literacy Scale in French University Students. Cronbach's alpha and McDonald's omega coefficients were 0.744 and 0.961, respectively, confirming an acceptable internal consistency (49).

As stated previously, when there are a small number of items in the scale (fewer than 10), Cronbach alpha values can be low. In this situation it may be better to calculate and report the **Average inter-item correlation** for the items. Average inter-item correlation is another way of analyzing internal consistency. It involves calculating the correlation coefficients between each pair of items within the questionnaire and then computing the average of these correlations. Optimal mean inter-item correlation values range from 0.2 to 0.4 (31). As observed in the study of the development of the European Portuguese version of the Psychometric Properties of the Sleep Locus of Control, an overall low Cronbach's alpha values in the questionnaire subscales (between 0.52 and 0.70) were observed. Therefore, the average of the inter-item correlations was assessed and revealed values between 0.18 to 0.32, which were indicative of a good internal consistency (50).

Another method for evaluating internal consistency is the evaluation of **Composite Reliability (CR)**. CR is based on the congeneric model that does not require equivalent factor loadings across items, which is defined as:

$$CR = \frac{\left(\sum \lambda_i\right)^2}{\left(\sum \lambda_i\right)^2 + \sum \left(1 - \lambda_i^2\right)},$$

where λ_i is the completely standardized factor loading (for which both indicators and latent constructs are standardized) of item i. CR values of 0.7 or higher denote good reliability (32).

The validation study of the Brazilian version of the Health Literacy Questionnaire serves as an example of a study that used both, Cronbach's alpha and composite reliability, to verify reliability. All nine scales of the questionnaire showed values above 0.76, except for one that was related to "Actively managing my health." The authors discussed the low internal consistency observed in this scale, which could be due to the fact that this construct is challenging for individuals to perform. Historically in Brazil, public policies and health practices have been centered on health professionals, which tends to minimize the individuals' participation in self-management of their health condition and the factors that determine it. Therefore, study participants may have found it difficult to answer about their ability to actively take care of their own health.

Test-Retest reliablity

Test-retest reliability measures the stability of the scores of a stable construct obtained from the same person on two or more separate occasions (51). Test-retest reliability can be evaluated using Pearson's product moment correlation coefficient (Pearson's r) or the intraclass correlation coefficient.

Pearson's r between the two questionnaires' responses can be referred to as the coefficient of stability. A higher coefficient indicates stronger test-retest reliability, which means that the questionnaire's measurement error is less likely to be caused by changes in the respondents' answers over time.

Test-retest reliability is not applicable for measuring transitory attributes like pain intensity and response to an medical intervention as changes in respondents' answers between assessments lead to instability in their responses (1).

It is also important o note the time lapsed between questionnaire administrations. The duration between the two administrations should not be too short because individuals may remember their responses, which may overestimate the test-retest reliability. Also, the duration between the two administrations should not be too long because individuals' responses may change due to other unknown factors. There is no consensus on the interval period, but most studies use an interval of 1 to 2 weeks (1,52).

Shakespear-Druery et al, 2022 did a test-retest reliability analysis on the Muscle-Strengthening Exercise Questionnaire (MSEQ). Participants completed the questionnaire in two different occasions, approximately 7 days apart. This study used both intraclass correlation coefficients (ICC) and Spearman's rank correlation coefficients for the test-retest reliability evaluation and overall, the MSEQ showed substantial test–retest reliability and adequate validity (53).

Inter-rater reliability

Inter-rater reliability is a critical measure in various research and assessment contexts, particularly when multiple raters or observers are involved in scoring or evaluating a given set of data or responses. It quantifies the degree of agreement or consistency among different raters in their judgments or assessments. This consistency can be estimated using the kappa statistic (coefficient k) and ICC. High inter-rater reliability values refer to a high degree of agreement between two examiners. Low inter-rater reliability values refer to a low degree of agreement between two examiners (54).

Some classifications have been proposed of the interpretation of coefficient k. k values falling within the range of 0.01 to 0.20 are often categorized as indicative of poor agreement, while those in the range of 0.21 to 0.40 are considered to demonstrate slight agreement. k values of 0.41 to 0.60 are typically characterized as reflecting fair agreement, while values in the range of 0.61 to 0.80 are associated with good agreement. k values within the range of 0.81 to 0.92, denoting very good agreement, while values between 0.93 and 1 are associated with excellent agreement (1).

Cunningham, 2015 conducted a pilot study evaluating the Inter-rater reliability of the Screenassist Lumbar Questionnaire. The inter-rater reliability of the SALQ was determined by performing ICCs for each of the questions using a 95% confidence interval. A Kappa analysis was performed for the overall scoring determination for referral recommendation (54).

3.3.2. Content validity

Ensuring that a questionnaire accurately represents the theoretical construct it is intended to assess is known as Content Validity. Content validity should be assessed once the initial version of the questionnaire is available. The process of content validation is particularly crucial in the development of a new questionnaire.

To validate the content of the questionnaire, a panel of experts who are familiar with the construct being measured should be assembled. These experts will evaluate whether the questionnaire questions adequately measure the intended construct, as well as whether the items are sufficient to measure the domain of interest.

Several approaches to quantify the judgment of content validity across experts are also available, such as the Content Validity Ratio (CVR) (1).

The CVR (content validity ratio) proposed by Lawshe (1975) (55) is a linear transformation of a proportional level of agreement on how many "experts" within a panel rate an item "essential", calculated in the following way:

$$\mathrm{CVR} = \frac{n_e - (\frac{N}{2})}{\frac{N}{2}}$$

where CVR is the content validity ratio, ne is the number of panel members indicating "essential," and N is the total number of panel members. The final evaluation to retain the item based on the CVR is depends on the number of panels.

The extent to which the construct of interest is operationalized is heavily dependent on how well the panel of experts can assess it. Therefore, selecting appropriate experts is essential to evaluate content validity adequately. Some of the example items that evaluate content validity include assessing if the questions were easy to understand, if they covered all problem areas related to pain, if the questionnaire can be used for future assessments and if the questionnaire lacks any vital questions regarding pain or if some questions violate privacy (25).

Chiwaridzo et al, 2017 performed a Content validity analysis of a low back pain questionnaire. Four content experts agreed to participate in the study and had to rate the relevance of each question/item in the

questionnaire on a scale of 1 to 4. Content Validity Ratio (CVR), was used to estimate quantitatively the content validity and results showed a CRF of 0.97, showing excellent content validity (56).

Face validity is another concept that relates to content validity. It is concerned with the extent to which respondents or laypersons judge the questionnaire items to be valid. This judgment is not based on the technical components of the questionnaire items but rather on whether the items seem to measure a meaningful construct for the respondents. Face validity may be the weakest way to establish the validity of a questionnaire, but it can motivate respondents to answer more truthfully. For instance, if patients perceive a quality of recovery questionnaire to evaluate how well they are recovering from surgery, they may be more likely to respond genuinely (1).

To assess face validity, a dichotomous scale can be used with the categorical options of "Yes" and "No". A "Yes" indicates a favorable item that is objectively structured and can be positively classified under its thematic category. A "No" indicates an unfavourable item. After collecting the data, Cohen's Kappa Index (CKI) is used to determine the face validity of the instrument. DM et al. (1975) recommended a minimum of 0.60 Kappa for inter-rater agreement (25).

3.3.3. Construct Validity

Ensuring that the items in a questionnaire accurately represent the underlying conceptual structure is known as construct validity.

Factor Analysis

Factor analysis is a statistical method used to identify the underlying domains or constructs within a developing measure. This approach is useful in establishing the construct validity of a measure. Factor analysis aims to identify the latent factors that influence observed variables or responses. These latent factors represent unobservable constructs or dimensions that can help simplify complex data and provide meaningful insights. In summary, the aim of factor analysis is to find the number and nature of common factors that explain the pattern of correlations between items. The common factors are unobserved (latent), thus factor analysis allows researchers to explore (exploratory factor analysis, EFA) or confirm (confirmatory factor analysis, CFA) the presence of underlying hypothesised constructs, measured by the corresponding items (39).

- **Exploratory Factor Analysis (EFA):**

  EFA can be a useful tool to examine patterns underlying a dataset. It helps researchers to understand how different items and constructs are related and can assist in developing new theories. EFA is suitable during early stages of instrument development. By utilizing EFA, the researcher can identify

items that do not empirically belong to the intended construct and should be removed from the survey (57).

The appropriate sample size needed for factor analysis is a complicated question, there are different ideas and several guiding rules of thumb in the literature. The questionnaire should be administered to a sample of sufficient size to allow exploratory factor analytic techniques to be performed. Ferguson and Cox (1993) suggest that 100 respondents is the absolute minimum number to be able to undertake this analysis. However, others would suggest that this is insufficient, and a rule of thumb would be five respondents per item (33). Recommendations varies between 100 and 1000 participants (58), but the general consensus is that larger sample sizes are generally better, as they will enhance the accuracy of all estimates and increase statistical power (57).

Prior to the extraction of the constructs, there are some tests which must be conducted to examine the adequacy of the sample and the suitability of data for FA. The most widely used method for assessing the sampling adequacy is the Kaiser–Meyer–Olkin (KMO), which compares simple correlations with partial correlations between variables. KMO values above 0.8 are considered good for factor analysis, and above 0.9 excellent. Bartlett's test of Sphericity (Bartlett 1950) provides a chi-square output that must be significant. It indicates the matrix is not an identity matrix and accordingly it should be significant ($p<0.05$) for factor analysis to be suitable. In brief, if the KMO indicates sample adequacy and Bartlett's test of sphericity indicates the item correlation matrix is not an identity matrix, then researchers can move forward with the FA (58).

Factor extraction can be conducted utilizing principal component analysis (PCA) or principal axis factoring, which are used most in research studies. In order to produce a more interpretable and simplified solution, rotation will help by maximizing high item loadings and minimizing low item loadings. Varimax rotation is the most common form of rotational methods for exploratory factor analysis and will often provide a simple structure (25,58) . Items loaded above 0.40, which is the minimum recommended value in research are considered for further analysis (25).

Cumulative Percentage of Variance, the Scree Plot criterion and Keiser's criterion are the most used criterion to choose the number of factors to be retained. The Cumulative Percentage of Variance states that in the natural sciences, factors should be stopped when at least 95% of the variance is explained although in the humanities, the explained variance is generally as low as 50–60% (58). The Scree Plot criterion states that the researcher should plot the eigenvalues of the correlation matrix in descending order, and then use a number of factors equal to the number of eigenvalues that occur prior to the last major drop in eigenvalue magnitude. The Kaiser criterion states that the researcher should use a number of factors equal to the number of the eigenvalues of the correlation matrix that are greater than one (59).

The last step of FAE is the interpretation where variables that belong to a particular construct are identified and named accordingly. This process is subjective, theoretical and inductive, as stated by Pett, Lackey et al. in 2003. It is significant that labels of constructs reflect the theoretical and conceptual intent (58).

- **Confirmatory factor analysis (CFA)**

CFA is a powerful statistical technique used to validate and confirm a hypothesized theoretical model based on collected data. It is especially useful when researchers have a clear understanding of the theoretical constructs they want to measure and have obtained prior evidence supporting the internal structure of the scale obtained in similar contexts (57).

To conduct a CFA, first, the factor model that needs to be tested needs to be precisely defined. This involves selecting the number of factors and defining the nature of the loadings between the factors and the measures. These loadings can be fixed at zero, fixed at another constant value, allowed to vary freely, or be allowed to vary under specified constraints (such as being equal to another loading in the model).

After collecting data, the correlations between each pair of measured variables need to be obtained. The researchers need to fit the model to the data. A method to obtain the estimates of factor loadings that were free to vary needs to be chosen. The most common model-fitting procedure is Maximum Likelihood Estimation. Asymptotically Distribution Free Estimation can be used in cases were data lacks multivariate normality.

Once the model is fitted to the data, the model's adequacy needs to be evaluated. The most commonly used test of model adequacy is the $\chi 2$ goodness-of-fit test. The null hypothesis for this test is that the model adequately accounts for the data, while the alternative is that there is a significant amount of discrepancy. This test is highly sensitive to sample size, such that tests involving large samples will generally lead to a rejection of the null hypothesis, even when the factor model is appropriate. An alternative is to use statistics that compare the fit of the proposed model to that of a "null model", such as Tucker-Lewis index, since these have been shown to be less sensitive to sample size.

Lastly, if the comparison of two models is intended, one of which is a reduced form of the other, the difference between their $\chi 2$ statistics can be examined, which will also have an approximately $\chi 2$ distribution. Almost all tests of individual factor loadings can be made as comparisons of full and reduced factor models. In cases where full and reduced models are not examined, the researcher can compare the Root mean square error of approximation (RMSEA), which is an estimate of discrepancy per degree of freedom in the model. (59)

Swami et al, 2020 performed an exploratory and confirmatory factorial analysis on a Bahasa Malaysia (Malay) translation of the Intuitive Eating Scale-2 (IES-2). In a sample of 519 individuals, the results found that when using Malay subsamples, the IES-2 scores reduced the number of factors, differing from the parent model. The confirmatory factor analysis did not confirm the parent 4-factor model, and the indices for the EFA-derived models were acceptable but not ideal. Out of the models tested, the EFA-derived 3-factor model had the best fit indices. The scores on this model had adequate internal consistency and were invariant across sex and ethnicity, however, the between-group differences in subscale scores were non-significant or negligible and evidence of the construct validity of Malay IES-2 scores was mixed. These results question the degree to with the construct evaluated by this questionnaire is applicable to Malasyan populations and non-Western populations (60).

Convergent and Divergent Validity

Convergent and divergent validity are integral aspects of assessing the construct validity of a questionnaire, ensuring that it accurately measures what it is intended to measure.

Convergent validity involves demonstrating that the questionnaire correlates positively with other measures that theoretically should be related. In essence, if the questionnaire is designed to capture a specific construct, it should exhibit a positive correlation with existing measures assessing the same or related constructs. This alignment reaffirms the questionnaire's ability to capture the intended construct effectively.

In opposition, **divergent validity** (also called discriminant validity) entails showcasing that the questionnaire correlates negatively or not at all with measures of unrelated constructs. This aspect is crucial in establishing that the questionnaire is distinct from other constructs and is not associated with unrelated variables.

By demonstrating strong convergent validity with related measures and robust divergent validity with unrelated ones, researchers can fortify the evidence supporting the construct validity of their questionnaire, ultimately contributing to the instrument's reliability and its capacity to yield meaningful and accurate results (1,33).

Wang et al, 2011 published a validation study of the Chinese version of the Walking Impeirment Questionaire (WIQ) in patients with peripheral arterial disease and type 2 diabetes mellitus. They assessed convergent and divergent validity, evaluating correlations between the WIQ and the SF-36 questionnaire using Spearman's rho. For this cohort of patients, it was expected that the three WIQ domains would show a high correlation with the domains related to physical health in the SF-36 (convergent validity) and show

a poor correlation with the domains concerning emotional aspects, especially the mental health domain of the SF-36 (divergent validity). High significant correlations between WIQ domains and SF-36 domains related to physical health were found, confirming convergent validity and poor significant correlation between the WIQ and the mental health domain of SF-36 were found, confirming divergent validity (61).

The **Average Variance Extracted** (AVE) is a statistical measure also commonly used to assess convergent validity. Fornell and Larcker (1981) (62) suggested that convergent validity is established when a latent construct accounts for no less than half of the variance in its associated indicators. They proposed using the average variance extracted (AVE) to represent the average amount of variance that a construct explains in its indicators relative to the overall variance of its indicators. For construct X, AVE is defned as follows:

$$AVE(X) = \frac{\sum_{i=1}^{p} \lambda_i^2}{\sum_{i=1}^{p} \lambda_i^2 + \sum_{i=1}^{p} Var(\varepsilon_i)} = \frac{1}{p}\left(\sum_{i=1}^{p} \lambda_i^2\right),$$

where p is the number of indicators of construct X, and λ_i is the completely standardized factor loading of the i^{th} indicator (both indicators and the construct are standardized). Thus, for construct X, the value of AVE is equivalent to the average of the square of completely standardized factor loadings across all its indicators. The AVE should not be lower than 0.5 to demonstrate an acceptable level of convergent validity, meaning that the latent construct explains no less than 50% of the indicator variance (32).
A validation study of the Korean version of the Toronto empathy questionnaire for the measurement of medical students' empathy demonstrated adequate convergent validity using the AVE statistical measure (63).

Known-groups Validity

Construct validity is the ability of an instrument to differentiate between different groups. Health constructs such as pain, mood, attitudes, disability, and quality of life are difficult to measure as they are not directly observable and have no standard method of measurement. To verify the construct validity of a test or questionnaire when there is no standard method, known-group validity is used. This involves administering the test or questionnaire to two groups with different levels of the construct to check if the hypothesized difference is reflected in the scores of both groups (34).
A validation study for the Spanish translation of the WIQ questionnaire conducted a known group validity and compared the walking impairment parameters for patients who reported that they did and did not (i) regularly exercise, (ii) have leg pain, and (iii) walk at least 10 minutes per day. In the three comparison groups, it was expected that patients who do regularly exercise, have no leg pain, and walk would have statistically better scores on each of the WIQ subscales than patients who do not regularly exercise, report

no leg pain, and do not walk, respectively. Results showed that For Spanish-speaking patients, there was no statistically significant difference for any walking parameter among exercisers vs. non-exercisers (64).

3.3.4. Criterion validity

Criterion-related validity is a measure of how adequately a test score can be used to estimate an individual's most likely standing on a specific measure of interest. Criterion-related validity comprises two types of validity evidence. Concurrent validity that measures the extent to which a test score is correlated with a criterion measure obtained simultaneously (concurrently), and Predictive validity that measures the degree to which a test score can forecast some criterion measure (65).

Concurrent Validity

Concurrent validity is a type of evidence that helps to support the use of a test for predicting other outcomes. It measures how well the results of a particular test or measurement correspond to those of a previously established measurement for the same construct. In brief, concurrent validity assesses how well the test can differentiate between groups that it should theoretically be able to distinguish between (25).

Predictive Validity

When a test accurately predicts what it is supposed to predict, it is considered to have predictively valid results. It can also refer to when scores from the predictor measure are taken first and then the criterion data is collected later. The best way to establish predictive validity is to perform a long-term validity study. Predictive validity studies take a considerable amount of time to complete and require a significant sample size to acquire meaningful data. In brief, predictive validity assesses the operationalization's ability to predict something it should theoretically be able to predict (65).

3.3.5. Responsiveness

Responsiveness is considered the longitudinal aspect of validity. When measuring change over time, it is important to determine whether the change in a person's questionnaire score is valid. This is called Responsiveness and it refers to the ability of a measurement instrument to accurately detect changes that occur over time in a given construct. The instrument should measure the right amount of change and not under- or overestimate it.

To assess responsiveness, the study design should be longitudinal with at least two measurements. The time points should be chosen in such a way that it can be expected that at least part of the study population will change in the construct of interest between the two time points.

When a gold standard is available, a correlation approach can be used to assess responsiveness. In this approach, we expect a strong correlation (>0.7) between the change score on the instrument of interest and the change score on the gold standard. However, even a correlation of 0.7 is quite challenging to achieve due to the measurement error included in both measurements.

In the absence of a gold standard, a construct approach can be used to test responsiveness. In this case, the hypotheses is about the magnitude of the change in score that can be expected after undergoing a treatment with known efficacy on the construct of interest. For example, if we know from previous research that a specific treatment will have a moderate effect on a specific outcome, we expect a medium effect size (between 0.3 and 0.5) using an instrument that measures the specific outcome in patients before and after receiving the treatment (36).

This way, Effect sizes (Cohen's criteria) can be calculated to quantify the magnitude of change. These effect sizes should be evaluated based on established benchmarks: a small effect size is indicated by d = 0.2, a medium effect size by d = 0.5, and a large effect size by d ≥ 0.8 (37).

Molen et al 2003, evaluated the responsiveness of a Clinical COPD Questionnaire (CCQ). The CCQ was readministered after a period of two months of successful smoking cessation in 36 subjects and responsiveness was tested using the Wilcoxon U test. Results showed that the CCQ significantly improved after two months smoking cessation showing responsiveness to change (66).

4. Conclusion

In this dissertation, a validation study of the Portuguese version of the VascuQoL-6 questionnaire was presented, where contributes to its validation as an instrument to assess the QoL of Portuguese patients with PAD were found.

Through the exploration of the validation process for the Portuguese version of the Vascular Quality of Life-6 Questionnaire, the complexities and nuances inherent in adapting research instruments to new cultural and linguistic contexts were unvailed and a overview analysis of the different validation methods was conducted.

Recognizing that no two clinical studies are alike, this dissertation emphasizes the need for adaptability and customization in the application of validation methods, aligning them with the unique characteristics of each investigation.

References

1. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Vol. 11, Saudi Journal of Anaesthesia. Medknow Publications; 2017. p. S80–9.
2. Shu J, Santulli G. Update on peripheral artery disease: Epidemiology and evidence-based facts. Atherosclerosis. 2018;275:379–81.
3. Aday AW, Matsushita K. Epidemiology of Peripheral Artery Disease and Polyvascular Disease. Circ Res. 2021;1818–32.
4. Conte SM, Vale PR. Peripheral Arterial Disease. Heart Lung Circ [Internet]. 2018;27(4):427–32. Available from: http://dx.doi.org/10.1016/j.hlc.2017.10.014
5. Van Den Houten MML, Gommans LNM, Van Der Wees PJ, Teijink JAW. Minimally Important Difference of the Absolute and Functional Claudication Distance in Patients with Intermittent Claudication. European Journal of Vascular and Endovascular Surgery. 2016;51:404–9.
6. Henni S, Ammi M, Semporé Y, Hersant J, Zegar G, Gourdier AS, et al. Treadmill Measured vs. Questionnaire Estimated Changes in Walking Ability in Patients With Peripheral Artery Disease. European Journal of Vascular and Endovascular Surgery. 2019;57(5):676–84.
7. Frans FA, Van Wijngaarden SE, Met R, Koelemay MJW. Validation of the Dutch version of the VascuQol questionnaire and the Amsterdam linear disability score in patients with intermittent claudication. Quality of Life Research. 2012;21(8):1487–93.

8. Nicolaï SPA, Kruidenier LM, Rouwet E V., Graffius K, Prins MH, Teijink JAW. The walking impairment questionnaire: An effective tool to assess the effect of treatment in patients with intermittent claudication. J Vasc Surg. 2009;50:89–94.
9. Nordanstig J, Karlsson J, Pettersson M, Wann-Hansson C. Psychometric properties of the disease-specific health-related quality of life instrument VascuQoL in a Swedish setting. Health Qual Life Outcomes. 2012;10:1–8.
10. Soria-juan B, Llanos L, Moral LR, Aparicio C, Gonzalez A, Shaw O, et al. Cross Cultural Adaptation and Validation of a Spanish Version of VascuQol-6 Questionnaire. Res Sq. 2021;1–17.
11. Poku E, Duncan R, Keetharuth A, Essat M, Phillips P, Woods HB, et al. Patient-reported outcome measures in patients with peripheral arterial disease: A systematic review of psychometric properties. Health Qual Life Outcomes [Internet]. 2016;14(1). Available from: http://dx.doi.org/10.1186/s12955-016-0563-y
12. Kumlien C, Nordanstig J, Lundström M, Pettersson M. Validity and test retest reliability of the vascular quality of life Questionnaire-6: A short form of a disease-specific health-related quality of life instrument for patients with peripheral arterial disease. Health Qual Life Outcomes. 2017;15(1):1–12.
13. Nordanstig J, Wann-Hansson C, Karlsson J, Lundström M, Pettersson M, Morgan MBF. Vascular Quality of Life Questionnaire-6 facilitates health-related quality of life assessment in peripheral arterial disease. J Vasc Surg. 2014;59:700–7.
14. Larsen ASF, Reiersen AT, Jacobsen MB, Kløw NE, Nordanstig J, Morgan M, et al. Validation of the Vascular quality of life questionnaire-6 for clinical use in patients with lower limb peripheral arterial disease. Health Qual Life Outcomes. 2017;15:184.
15. Correia M, Andrade-Lima A, Oliveira PL, Domiciano RM, Domingues WJ, Wolosker N, et al. Translation and Validation of the Brazilian-Portuguese Short Version of Vascular Quality of Life Questionnaire in Peripheral Artery Disease Patients with Intermittent Claudication Symptoms. Ann Vasc Surg. 2018;51:48–54.
16. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. BMC Med Res Methodol. 2010;10.
17. Wild D, Grove A, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value in Health [Internet]. 2005;8(2):94–104. Available from: http://www.ispor.org
18. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery

disease: Executive Summary: A report of the American college of cardiology/American Heart Association task force on clinical practice guidelines. Circulation. 2017. 686–725 p.

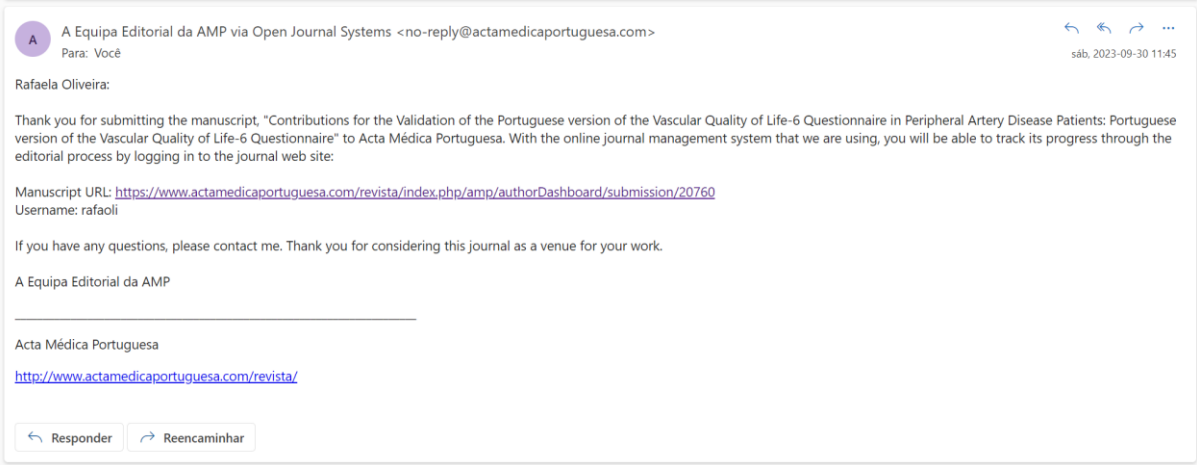
19. Kruidenier LM, Nicolaï SPA, Willigendael EM, de Bie RA, Prins MH, Teijink JAW. Functional claudication distance: A reliable and valid measurement to assess functional limitation in patients with intermittent claudication. BMC Cardiovasc Disord. 2009;9:1–7.
20. Treat-Jacobson D, McDermott MM, Bronas UG, Campia U, Collins TC, Criqui MH, et al. Optimal Exercise Programs for Patients with Peripheral Artery Disease. Circulation. 2019;139:10–33.
21. Gardner A, Skiner J. Progressive vs single-stage treadmill tests for evaluation of claudication. Med Sci Sports Exerc. 1991;23:402–8.
22. Ferreira PL. Development of the Portuguese version of MOS SF-36. Part II --Validation tests. Acta Med Port. 2000;13:119–27.
23. Hurtig-Wennlf A, Hagstrmer M, Olsson LA. The International Physical Activity Questionnaire modified for the elderly: Aspects of validity and feasibility. Public Health Nutr. 2010;13:1847–54.
24. Pedras S, Oliveira R, Veiga C, Silva I. What Do Patients Know About Peripheral Arterial Disease? A Knowledge Questionnaire. Portuguese Journal of Cardiac Thoracic and Vascular Surgery. 2023;29(4).
25. Hamed Taherdoost A, Lumpur K. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research [Internet]. Vol. 5, International Journal of Academic Research in Management (IJARM). 2016. Available from: https://ssrn.com/abstract=3205040
26. Hair J, Back W, Aderson R. Multivariate Data Analysis. 7th ed. Pearson Prentice Hall; 2010.
27. McHorney C, Ware J, Lu J, Sherbourne C. The MOS 36-item Short-Form Healthy Survey (SF-36): III. Test of data quality, scaling assumptions and reliability across diverse patient groups. Med Care. 1994;32(1):40–66.
28. Bland JM, Altman D. Statistics notes: Cronbach's alpha. BJM. 1997;314:572.
29. Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. J Pers Assess. 2003;80(1):99–103.
30. Stensen K, Lydersen S. Internal consistency: from alpha to omega? Tidsskrift for den Norske Laegeforening. 2022;142(12).
31. Pallant J. SPSS Survival Manual: A Step By Step Guide to Data Analysis Using SPSS Program. 6th edition. London: McGraw-Hill Education; 2016.
32. Cheung GW, Cooper-Thomas HD, Lau RS, Wang LC. Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. Asia Pacific Journal of Management. 2023;

33. Rattray J, Jones MC. Essential elements of questionnaire design and development. Vol. 16, Journal of Clinical Nursing. 2007. p. 234–43.
34. Davidson M. Known-Groups Validity. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research [Internet]. Dordrecht: Springer Netherlands; 2014. p. 3481–2. Available from: https://doi.org/10.1007/978-94-007-0753-5_1581
35. Hardman RL, Jazaeri O, Yi J, Smith M, Gupta R. Overview of classification systems in peripheral artery disease. Vol. 31, Seminars in Interventional Radiology. Thieme Medical Publishers, Inc.; 2014. p. 378–88.
36. Mokkink L, Terwee C, de Vet H. Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. J Clin Epidemiol. 2021 Dec 1;140:159–62.
37. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough . J Grad Med Educ. 2012 Sep 1;4(3):279–82.
38. Piedmont R. Encyclopedia of Quality of Life and Well-Being Research. 1st ed. Springer Dordrecht; 2014. 3303–3304 p.
39. Yusoff MSB, Arifin WN, Hadie SNH. ABC of questionnaire development and validation for survey research. Vol. 13, Education in Medicine Journal. Penerbit Universiti Sains Malaysia; 2021. p. 97–108.
40. Hamed Taherdoost A, Lumpur K. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research [Internet]. Vol. 5, International Journal of Academic Research in Management (IJARM). 2016. Available from: https://ssrn.com/abstract=3205040
41. Khalid A. Questionnaire designing and validation [Internet]. Vol. 62, Journal of the Pakistan Medical Association. 2012. Available from: http://ecommons.aku.edu/
42. Rattray J, Johnston M, Wildsmith JAW. The intensive care experience: development of the ICE questionnaire. J Adv Nurs [Internet]. 2004 Jul 1;47(1):64–73. Available from: https://doi.org/10.1111/j.1365-2648.2004.03066.x
43. Borba L de O, Capistrano FC, Ferreira ACZ, Kalinke LP, Mantovani M de F, Maftum MA. Adaptation and validation of the Measuring of Treatment Adherence for mental health. Rev Bras Enferm. 2018;71:2243–50.
44. Wiklund I. The Nottingham Health Profile--a measure of health-related quality of life. Scand J Prim Health Care Suppl [Internet]. 1990;1:15–8. Available from: http://europepmc.org/abstract/MED/2100359
45. Wallace M. Katz Index of Independence in Activities of Daily Living (ADL) [Internet]. Vol. 2, Issue Number. 2007. Available from: http://www.nursingcenter.com

46. Hinkin TR. A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. Organ Res Methods [Internet]. 1998 Jan 1;1(1):104–21. Available from: https://doi.org/10.1177/109442819800100106
47. Walde P, Völlm BA. The TRAPD approach as a method for questionnaire translation. Front Psychiatry. 2023;14.
48. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Vol. 2, International journal of medical education. 2011. p. 53–5.
49. Montagni I, González Caballero JL. Validation of the Mental Health Literacy Scale in French University Students. Behavioral Sciences. 2022 Aug 1;12(8).
50. Vieira P, Silva J, Gomes AA, Vincent N, de Azevedo MHP, Marques DR. Psychometric Properties of the Sleep Locus of Control (SLOC) Scale in a Portuguese Sample. Journal of Rational – Emotive and Cognitive – Behavior Therapy. 2023 Mar 1;41(1):193–208.
51. Vilagut G. Test-Retest Reliability. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research [Internet]. Dordrecht: Springer Netherlands; 2014. p. 6622–5. Available from: https://doi.org/10.1007/978-94-007-0753-5_3001
52. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003 Aug 1;56(8):730–5.
53. Shakespear-Druery J, De Cocker K, Biddle SJH, Bennie J. Muscle-Strengthening Exercise Questionnaire (MSEQ): An assessment of concurrent validity and test-retest reliability. BMJ Open Sport Exerc Med. 2022 Feb 14;8(1).
54. Cunningham S. Inter-rater reliability and validity of the screenassist lumbar questionnaire: A pilot study. Journal of Manual and Manipulative Therapy. 2015 Dec 1;23(5):254–62.
55. Lawshe CH. A Quantitative Approach To Content Validity. Vol. 28. 1975.
56. Chiwaridzo M, Chikasha TN, Naidoo N, Dambi JM, Tadyanemhandu C, Munambah N, et al. Content validity and test-retest reliability of a low back pain questionnaire in Zimbabwean adolescents. Arch Physiother. 2017 Dec 1;7(1).
57. Knekta E, Runyon C, Eddy S. One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sci Educ. 2019 Mar 1;18(1).
58. Taherdoost H, Sahibuddin S, Jalaliyoon N. Exploratory Factor Analysis; Concepts and Theory. In: Advances in Applied and Pure Mathematics. 2014. p. 375–82.
59. Decoster J. Overview of Factor Analysis [Internet]. 1998. Available from: http://www.stat-help.com
60. Swami V, Todd J, Zahari HS, Nor NA, Toh EKL, Barron D. Dimensional structure, psychometric properties, and sex and ethnic invariance of a Bahasa Malaysia (Malay) translation of the Intuitive Eating Scale-2 (IES-2). Body Image. 2020 Mar 1;32:167–79.

61. Jie Wang, Yan Cui, Bian RW, Mo YZ, Haidi Wu, Ling Chen. Validation of the Chinese version of the Walking Impairment Questionnaire in patients with both peripheral arterial disease and type 2 diabetes mellitus. Diab Vasc Dis Res. 2011 Jan;8(1):29–34.
62. Fornell C, Larcker D. Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research. 1981;18(1):39–50.
63. Yeo S, Kim KJ. A validation study of the Korean version of the Toronto empathy questionnaire for the measurement of medical students' empathy. BMC Med Educ. 2021 Dec 1;21(1).
64. Collins TC, Suarez-Almazor M, Petersen NJ, O'Malley KJ. A Spanish translation of the Walking Impairment Questionnaire was validated for patients with peripheral arterial disease. J Clin Epidemiol. 2004 Dec;57(12):1305–15.
65. Primis H. Psychological Testing and Assessment: An Introduction to Tests and Measurement, 7th Edition [Internet]. 1976. Available from: http://www.primisonline.com
66. Van Der Molen T, Willemse BW, Schokker S, Ht Ten Hacken N, Postma DS, Juniper EF. Health and Quality of Life Outcomes Development, validity and responsiveness of the Clinical COPD Questionnaire [Internet]. 2003. Available from: http://www.hqlo.com/content/1/1/13

Annex I



A Equipa Editorial da AMP via Open Journal Systems <no-reply@actamedicaportuguesa.com>
Para: Você
sáb, 2023-09-30 11:45

Rafaela Oliveira:

Thank you for submitting the manuscript, "Contributions for the Validation of the Portuguese version of the Vascular Quality of Life-6 Questionnaire in Peripheral Artery Disease Patients: Portuguese version of the Vascular Quality of Life-6 Questionnaire" to Acta Médica Portuguesa. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Manuscript URL: https://www.actamedicaportuguesa.com/revista/index.php/amp/authorDashboard/submission/20760
Username: rafaoli

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

A Equipa Editorial da AMP

Acta Médica Portuguesa

http://www.actamedicaportuguesa.com/revista/

Responder Reencaminhar