International Journal of Cerebrovascular Disease and Stroke
Themistocleous IC et al. Int J Cerebrovasc Dis Stroke 6: 163.
www.doi.org/10.29011/2688-8734.000063
www.gavinpublishers.com

Research Article

Reliability and Validity of the Greek Version of Sickness Impact Profile Questionnaire

Irene-Chrysovalanto Themistocleous[1*], Lampropoulou Sofia[2], Papadopoulou-King Christina[1], Stefanakis Manos[1]

[1]School of Life and Health Sciences, Department of Health Science, Program of Physiotherapy, University of Nicosia, Cyprus

[2]Physiotherapy Department, School of Health Rehabilitation Sciences, University of Patras, Greece

***Corresponding author:** Irene-Chrysovalanto Themistocleous, School of Life and Health Sciences, Department of Health Science, Program of Physiotherapy, University of Nicosia, Cyprus

Citation: Themistocleous IC, Lampropoulou S, Papadopoulou-King C, Stefanakis M (2023) Reliability and Validity of the Greek Version of Sickness Impact Profile Questionnaire. Int J Cerebrovasc Dis Stroke 6: 163. DOI: https://doi.org/10.29011/2688-8734.100163

Received Date: 27 September, 2023**; Accepted Date:** 02 October, 2023**; Published Date:** 05 October, 2023

Abstract

Background: The objectives were to assess the validity and reliability of the Greek version of the Sickness Impact Profile (SIP-GR) questionnaire. **Methods:** SIP-GR was tested for test-retest reliability, internal consistency and validity in 90 participants (54.4% males and 45.6% females) with obesity, cardiac, pulmonary and musculoskeletal problems. The questionnaire was administrated twice by one examiner, within an interval of 1-week. During this period, participants with cardiac, pulmonary and musculoskeletal problems underwent 2-weekly physiotherapy sessions. Treatment related effects were considered in the analysis. **Results:** SIP-GR demonstrated an excellent internal consistency. The overall Cronbach's alpha for SIP-Total score >0.9, for SIP-Psychological was >0.8 and for SIP-Physical >0.9. The categories scores were all >0.5 except for Communication and Work category. Test-retest reliability for the total score was ICC=0.691 for all subjects, 0.562 for those that reported a subjective change in their health status due to treatment and 0.999 for those that reported no change in health status. Similar results were found for the Physical and Psychological component. Strong negative correlation was found between SIP-Gr total score and Short Form Health Survey SF-36 (SF-36) total score (r=-0.66), physical component of SIP-GR and Physical Health of SF-36 (r=-0.62) and between psychological component of SIP-GR and Mental health of SF-36 (r=-0.61) at the initial assessment. At reassessment the same correlations were moderate due to different treatment effects in the scores of the two questionnaires. Finally, minimum detectable change (MDC) was 4.6-5.5 points for the overall score, 6.4-7 for Physical component and 3.5-4.9 for psychosocial component, at initial assessment and re-assessment respectively. **Conclusion:** SIP-GR has shown to be valid and reliable for the assessment of patients with cardiac, pulmonary, musculoskeletal diagnosis and obesity. Further studies should assess its ability to identify clinically meaningful changes.

Keywords: Sickness Impact Profile; Questionnaires; Validation; Reliability

Introduction

Health-related quality of life (HRQOL) is important in a variety of diseases and one of the most important outcome measures after an intervention. Generic questionnaires are important as they measure all the aspects of the disease and reflect the overall impact of the disease and the benefits of treatment [1].

The Sickness Impact Profile (SIP) is a generic questionnaire of health-related functional status [2] and can be used across different types and severities of diseases [3]. SIP scores are available for approximately 18 different diseases or populations [4]. In order to measure the health status, participants must be questioned in their own language [5].

SIP has been translated into several languages including Arabic, Chinese for Hong-Kong, Danish, Dutch, Dutch for Belgium, English for Mexico, English for the UK, Finnish, French, French for Belgium, Italian, German, Norwegian, Portuguese, Romanian, Russian, Spanish, Spanish for Mexico, Spanish for the USA, Tamil and Thai, Swedish [6]. The aim of this study was the cross-cultural adaptation of SIP in Greek language (SIP-GR) and the assessment of its psychometric properties.

Materials and Methods

The study was approved by the Cyprus Bioethics Committee (ΕΕΒΚ ΕΠ 2018.01.148) and whole adaptation process was approved by the SIP developers.

Sample size calculation

Sample size calculations were based on the intraclass correlation coefficient (ICC) of test-retest reliability of total scores. Using an acceptable ICC of at least 0.7, an expected ICC of at least 0.9 a power of 80% a significance level of 0.05 the sample size required is 19 subjects [7]. An additional twenty percent of drop-out rate was included bringing the total number of subjects required to 21. We recruited ninety subjects all together, sixty patients with cardiopulmonary and musculoskeletal diseases and thirty apparently healthy but obese or overweight subjects (at risk group).

Participants

Ninety (n=90) participants were included in this study after signing a written informed consent. Recruitment was achieved via advertisement in the local area of Nicosia, Cyprus and from the patient list of a local Physiotherapy clinic. Participants were invited to participate in this study if they met the following inclusion criteria: ≥18 years old, comprehension of Greek language, obese and overweight (≥25 body mass index), had recently undergone cardiac surgery (within 2-months after surgery), suffer from primary osteoarthritis of the hip or knee joint with symptoms lasting at least 2-months, complained for nonspecific low back pain lasting at least 2-months or were diagnosed with pulmonary diseases with recurrent exacerbations. All the participants (n=60) except overweight and obese, attended 2 weekly conservative physiotherapy sessions according to their diagnosis, during the period of data collection. Participants reported information on sociodemographic variables such as gender, age and occupation. In order to examine the test-retest reliability, all subjects were asked to fill the questionnaire twice at one-week interval. In order to assess its validity, SIP was compared to 36-item Short Form Health Survey (SF-36).

Cross-cultural adaptation process

In general, cultural adaptation of SIP followed the internationally accepted guidelines for cultural adaptation of patient reported outcomes published by Beaton et al. [8]. The process intended to produce equivalency between the original and translated versions in terms of content and to adapt it culturally so that the original meaning and intent of the items are maintained. The process involved the following 6 steps:

Step 1: Forward translation and harmonization of forward translation

Two of the authors, both bilingual physiotherapists, native in Greek language independently translated the original version of SIP into Greek and produced a report. The two versions were then synthesized into one initial translation by consensus of the two reviewers and a new report was produced which was send to the SIP developers.

Step 2: Backward translation and harmonization of backward translation

Two bilingual translators one physiotherapist and one language expert, native in English language and fluent in Greek, produced two independent back translations of the original Greek version. All translators were kept blind to the original English version of SIP.

An expert committee consisting of the translators, one more Physiotherapist and one medical doctor produced the pre-final version of the SIP-GR. The committee made a great effort to ensure semantic, idiomatic, experimental and conceptual equivalence existed between the original and the SIP-GR versions. The whole process was documented and a report was again sent to SIP developer.

Step 3: Validation of the pre final version

Validation of the translation was performed by evaluating the comparability of language and the similarity of interpretation using a 7-point Likert scale ranging from 1 (extremely comparable/similar) to 7 (not at all comparable/not at all similar). 10 bilingual individuals independently compared the English and the translated versions item by item and rated each one in terms of comparability and similarity. Any item above 3 in comparability or 2.5 in similarity was deemed appropriate for revision. No item needed revision. The same 10 individuals were used for cognitive debriefing. Cognitive equivalence of the translated version was tested between various educational backgrounds and different regions of Greece in order to capture differences in dialects among individuals.

Step 4: Expert committee review

The committee assessed all the reports from the previous steps and made all the necessary modifications to optimize the final version.

Step 5: Proofreading

A proofreading company was consulted to correct the final version in terms of spelling, diacritical, grammatical, or any other errors. Following this step, the Final version was ready for pilot testing.

Step 6: Pretesting (pilot study)

The pilot testing of the questionnaire was performed in 10 individuals (7 males, 3 females) from the general population approached randomly in a local mall. Individuals had to be overweight or obese and be able to speak, read and understand Greek. The 10 participants (mean age 47.7 ± 22.64 years) completed the self-administrated SIP-136 Greek version of the questionnaire. Eight participants were obese, 2 were overweight, 7 had hypertension, but only 5 of them were taking medication, 4 participants had osteoarthritis (OA), 1 was a stroke survivor, 1 was diagnosed with Parkinson's disease (PD), 1 with asthma and 1 had a stent due to a congenital heart disease. All individuals were asked to complete the questionnaire without the help of an interviewer and record any difficulties in comprehending any item on a standard form. SIP scores and the time needed to fill the questionnaire were also recorded. None of the participants recorded any problem with comprehending any item and none left any item unanswered. The mean time to complete the questionnaire was 34.2 ± 7.32 minutes and the mean total score was 14.8 ± 9.37 (Physical component = 14.6 ± 16.43, Psychosocial component = 12.5 ± 7.88).

Measurement instruments

Sickness Impact Profile

SIP is a generic questionnaire designed to subjectively assess the physical and psychological functioning in a wide range of diseases. SIP consists of 136-items that are divided into 12 categories related to daily living which are then grouped into physical and psychosocial dimensions [9,10]. Physical domain includes: ambulation, mobility, body care/movement. Psychosocial domain includes: social interaction, communication, alertness behaviour, emotional behaviour. Sleep and rest, eating, home management, recreation and pastimes, and work are considered independent categories. Each item is a question in present tense and patients is asked to reply how they feel at the time of test administration. Replies in all questions are in binary form (´´Yes/No'') and the patient selects all the questions that are applicable to them. The total score is calculated by summing up the domain scores and the result is expressed as a percentage of the maximum possible score based on the answers (0-100). A higher score represents a more severe impact of the disease on health.

36-item Short Form Health Survey

The 36-item Short Form Health Survey SF-36 (SF-36) is a generic questionnaire which includes 36 questions based on the general health status, divided into 8 subscales which ultimately provide two scores of physical and mental health status [11,12]. SF-36 subscale scores, range between 0 and 100, where greater score shows better HRQOL [13].

Assessment Analysis

Reliability

Test-retest reliability

The two-way mixed model, intraclass correlation coefficient (ICC), with absolute agreement was used to assess the reliability between the first and the second time point in the domain and overall scores of SIP-GR (1-week interval). If ICC values were ≤ 0.40 reliability was considered poor, between 0.40-0.75 moderate, between 0.75 and 0.90 substantial and > 0.90 excellent [14]. Because most subjects received treatment during the one-week test/retest period, all participants were asked to subjectively rate their change of health status during the last week using a 0-100% scale. Test-retest reliability was calculated separately on those subjects that rated their health status change as zero (n=36). Those patients were all the obese (no treatment), 5 musculoskeletal and 1 Cardiopulmonary patient. Moreover, the standard error of measurement (SEM) was calculated based on the formula SEM=SD□√(1−ICC), where SD is the standard deviation of the initial assessment, and ICC is the value obtained by the analysis of test-retest. In addition, minimal detectable change at the 90% confidence level (MDC) was calculated based on the MDC= 1,65 x √2 x SEM formula [15]. Lower SEM values indicates better reliability of the measure, whereas lower MDC values indicates a more sensitive measure [16].

Internal consistency

Internal consistency was determined using the Cronbach's alpha. Values higher than 0.7 are considered as sufficient [14].

Validity

Construct validity

Spearman's rank correlation coefficient (rs) was used to test the construct validity between SIP-GR, and SF-36 questionnaire, since the results were not normally distributed. Rs values of 0.00-0.30 are considered weak, 0.31-0.59 moderate and 0.60-1.00 strong [17]. The hypothesis is that there will be significant correlation between the overall scores of the two questionnaires and that the correlation of the psychological component of SIP will be higher with the Mental health of SF-36 than the Physical health and vice versa.

Known group validity

The total score and as well as the domain scores were compared between the three groups of patients namely the obese, the musculoskeletal and the cardiorespiratory group. The hypothesis is that at risk group (obese) will be significant different than the two patient groups.

Data Analysis

Data analysis was performed using SPSS (Version 25.0) and Jamovi (Version 2.2.5). The level of significance was set at 0.05. Descriptive statistics were reported using means and standard deviations (SD), or frequencies for the demographic characteristics. SIP scores are summarized for each of the 12 categories and for the domain and total score. Kruskal-Wallis test was used to assess differences between the three patient groups. Pre and post comparisons were performed to determine the changes in each questionnaire between initial assessment and re-assessment using Friedman's Test. Test-retest reliability was assessed using ICC and 95% confidence intervals. Internal consistency was evaluated using Cronbach's alpha. Construct validity was evaluated via Spearman's rank correlation coefficient.

Results

A total of 90 participants included in this study. Thirty of the participants had a history of musculoskeletal problems, 30 participants had a history of cardiac or pulmonary problems and the remaining 30 participants were obese or overweight. There was a good balance between male (54.4%) and female (45.6%) gender. Table 1 summarizes the characteristics of the respondents.

Variable **Mean ± SD**	***Total participants (n=90)***
Age (years)	51.5±19.18
BMI (Kg/m^2)	31.6±4.58
Completion time (minutes)	34.7±6.56
Number and percentage (%)	
Gender	
Male	49 (54.4%)
Female	41 (45.6%)
Health Condition	
Overweight	4 (4.4%)
Obese	26 (28.9%)
OSAS	3 (3.3%)
Lumbar Pain	14 (15.6%)
Knee pain	16 (17.7%)
Heart failure	5 (5.6%)
COPD	5 (5.6%)
CABG	7 (7.8%)
AVR	5 (5.6%)
Asthma	5 (5.6%)
Occupation	
Administrative work	16 (17.8%)
Business management	4 (4.4%)
College students	7 (7.8%)
Teaching personnel	7 (7.8%)
Pensioneer	27 (30%)
Health care professionals	4 (4.4%)

Manual workers	7 (7.8%)
Engineers	10 (11.1%)
Arts	4 (4.4%)
Military officers	1 (1.1%)
Hairdressers	2(2.2%)
Professional drivers	1 (1.1%)
Abbreviations: BMI: Body Mass Index**; OSAS:** Obstructive Sleep Apnoea Syndrome**; CABG:** Coronary Artery Bypass Graft**; AVR:** Aortic Valve Replacement; **COPD:** Chronic Obstructive Pulmonary Disease.	

Table 1: Respondents' characteristics.

Table 2 presents the means and SDs of the different categories, domains and total score of SIP for the three groups of patients at initial assessment and re-assessment. Generally, the scores were lower for obese subjects and higher for Cardiopulmonary patients. Musculoskeletal patients were lower than Cardiopulmonary group and higher than Obese group in most categories at initial assessment. However, as most of them improved with treatment they achieve lower scores than Obese in several categories, in Physical component and total score during re-assessment (Table 2).

		Cardiopulmonary	***Musculoskeletal***	***Obese***
Sleep and Rest	**Initial**	36.0 (19.40)	26.4 (13.70)	3.7 (7.95)
	Re-assessment	16.3 (11.50)	10.8 (9.58)	4.5 (8.14)
Emotional Behavior	**Initial**	13.1 (16.30)	14.9 (8.72)	7.2 (13.80)
	Re-assessment	8.4 (10.80)	8.8 (6.17)	7.7 (14.00)
Body Care and Movement	**Initial**	23.1 (17.70)	17.2 (13.90)	11.4 (7.91)
	Re-assessment	19.0 (18.80)	5.9 (5.67)	11.5 (7.84)
Home Management	**Initial**	54.8 (28.50)	25.8 (14.70)	17.4 (12.60)
	Re-assessment	47.8 (31.20)	15.3 (12.50)	17.4 (12.60)
Mobility	**Initial**	33.7 (22.40)	17.4 (15.80)	2.3 (6.34)
	Re-assessment	18.8 (16.30)	7.4 (6.77)	2.3 (6.34)
Social Interaction	**Initial**	29.5 (16.00)	13.3 (10.60)	3.5 (9.88)
	Re-assessment	16.3 (14.30)	8.2 (7.01)	3.8 (9.85)
Ambulation	**Initial**	29.0 (8.14)	27.6 (18.00)	19.1 (7.56)
	Re-assessment	28.3 (16.40)	21.4 (19.20)	19.1 (7.56)
Alertness Behavior	**Initial**	1.0 (2.75)	1.1 (3.55)	2.2 (7.17)
	Re-assessment	0.6 (2.20)	0.6 (2.33)	2.2 (7.17)
Communication	**Initial**	4.2 (17.70)	3.1 (6.58)	0.3 (1.61)
	Re-assessment	1.0 (3.04)	1.1 (3.46)	0.3 (1.61)
Work	**Initial**	33.6 (32.60)	25.5 (23.60)	9.3 (9.83)
	Re-assessment	25.4 (30.20)	12.8 (17.30)	8.8 (9.54)
Recreation and Pastimes	**Initial**	34.9 (12.50)	29.0 (13.50)	21.4 (11.90)
	Re-assessment	27.7 (11.80)	17.4 (11.00)	22.3 (11.70)
Eating	**Initial**	10.4 (9.22)	1.0 (2.99)	3.1 (5.37)
	Re-assessment	9.5 (10.10)	0.8 (2.52)	3.1 (5.37)
Physical Component	**Initial**	26.5 (13.00)	18.3 (12.40)	11.4 (6.79)
	Re-assessment	21.2 (15.30)	7.3 (5.01)	11.4 (6.78)

Psychosocial Component	**Initial**	15.2 (8.90)	9.0 (5.79)	3.3 (6.97)
	Re-assessment	8.4 (6.81)	5.3 (3.79)	3.5 (6.94)
Total Score	**Initial**	24.2 (9.11)	15.4 (6.71)	9.1 (7.32)
	Re-assessment	17.7 (9.34)	7.5 (3.50)	9.3 (7.28)

Table 2: SIP categories, domain and total scores by patient group (Mean ± SD).

Reliability

Internal consistency

Cronbach's alpha was calculated for the individual categories and the domain and the total scores of SIP. The overall Cronbach's alpha for SIP-Total score >0.9, for SIP-Psychological was >0.8 and for SIP-Physical >0.9. The categories scores were >0.5 except for Communication and Work (Table 3).

	Internal consistency (Cronbach's α)	
	Initial Assessment	**Re-assessment**
Sleep and Rest	0.755	0.593
Emotional Behavior	0.693	0.620
Body Care and Movement	0.887	0.894
Home Management	0.896	0.898
Mobility	0.805	0.729
Social Interaction	0.881	0.853
Abulation	0.807	0.836
Alertness Behavior	0.508	0.593
Communication	0.063	-0.017
Work	0.218	0.177
Recreation and Pastimes	0.636	0.614
Eating	0.645	0.679
Physical Component	0.936	0.937
Psychosocial Component	0.871	0.870
Total Score	0.965	0.962

Table 3: Cronbach's alpha for categories, domain and total score of SIP.

Test-retest reliability

Intraclass correlation coefficient (ICC) was higher in the subjects that had no change in their health status compared to those that changed due to treatment. ICC for the total score was 0.691 for all subjects, 0.562 for those that reported a subjective change in their health status and 0.999 for those that reported no change in health status. This means there was moderate test-retest reliability for those that had a treatment effect and excellent reliability for those that had no change in health status. The results are similar for the Physical and Psychosocial domains and are included in (Table 4).

	Physical component	***Psychosocial component***	***Total score***
No change in Health status (N=36)	1 (0.999-1.000)	0.996 (0.974-1.000)	0.999 (0.993-1.000)
Change in health status (n=54)	0.516 (0.364-0.654)	0.458 (0.254-0.667)	0.562 (0.397-0.683)
All subjects (N=90)	0.599 (0.460-0.718)	0.663 (0.495-0.807)	0.691 (0.565-0.783)

Table 4: SIP Test-retest reliability (ICC, 95% CI).

The effect of treatment in the questionnaire scores and the reliability of the instrument are supported by the comparisons of pre and post values among patients with significant improvement (N=54) and those without improvement (N=36). Only those with no change in health status had no significant difference between the pre and post values in both questionnaires.

		All subjects		*N=36*		*N=54*	
SIP Physical component	**Initial assessment**	15.0 (16.78)	p=0.001	10.0 (5.80)	p=ns	22.1 (19.52)	p=0.001
	Re-assessment	9.8 (9.02)		10.0 (5.80)		8.8 (15.51)	
SIP Psychosocial component	**Initial assessment**	8.1 (12.95)	p=0.001	1.3 (3.59)	p=ns	12.1 (9.31)	p=0.001
	Re-assessment	4.4 (6.08)		1.4 (4.11)		6.0 (4.97)	
SIP Total score	**Initial assessment**	14.1 (15.96)	p=0.001	7.1 (5.53)	p=ns	18.9 (12.91)	p=0.001
	Re-assessment	8.4 (7.26)		7.3 (5.50)		9.2 (11.88)	
SF36 Physical Health	**Initial assessment**	36.8 (23.30)	p=0.001	52.2 (15.20)	p=ns	31 (14.00)	p=0.001
	Re-assessment	45.6 (19.00)		51.6 (21.65)		41.2 (17.60)	
SF36 Mental Health	**Initial assessment**	53.1 (25.69)	p=0.001	73.7 (10.67)	p=ns	48.2 (10.43)	p=0.001
	Re-assessment	65.7 (20.85)		73.1 (10.45)		59.1 (18.38)	
SF36 Total score	**Initial assessment**	43.3 (28.29)	p=0.001	63.9 (13.67)	p=ns	35.5 (11.22)	p=0.001
	Re-assessment	54.6 (20.95)		64.0 (15.32)		48.2 (16.38)	

Table 5: Comparison between initial assessment and re-assessment among those with health improvement and those without improvement (Median/IQR).

SIP-GR total score demonstrated an SEM of 5.5 and MDC of 12.8 points for the assessments at baseline, and an SEM of 4.6 and MDC of 10.8 points for the re-assessment using the ICC value of all subjects (0.691).

Construct Validity

There were strong negative correlations between SIP Physical component and SF36 Physical Health (rs=-0.621, p=0.001), between SIP Psychosocial component and SF36 Mental Health (rs=-0.619, p=0.001) and between SIP Total score and SF36 Total score (rs=-0.661, p=0.001) at the initial assessment (Table 6). Those correlations were still significant but only moderate (rs=-0.455, p=0.001, rs=-0.437, p=0.001, rs=-0.341, p=0.001 respectively) at reassessment (Table 7).

	SF36 Physical Health	*SF36 Mental Health*	*SF36 Total Score*
SIP Physical Domain	rs=-0.621, p=0.001	rs=-0.424, p=0.001	rs=-0.540, p=0.001
SIP Psychosocial Domain	rs=-0.591, p=0.001	rs=-0.619, p=0.001	rs=-0.603, p=0.001
SIP Total Score	rs=-0.703, p=0.001	rs=-0.591, p=0.001	rs=-0.661, p=0.001
Abbreviations: SIP: Sickness Impact Profile questionnaire**; SF-36:** The 36-Item Short Form Survey**.**			

Table 6: Spearman's Correlation between SIP and SF36 at initial assessment (N=90).

	SF36 Physical Health	*SF36 Mental Health*	*SF36 Total Score*
SIP Physical Domain	rs=-0.455, p=0.001	rs=-0.050, p=0.639	rs=-0.253, p=0.016
SIP Psychosocial Domain	rs=-0.401, p=0.001	rs=-0.437, p=0.001	rs=-0.403, p=0.001
SIP Total Score	rs=-0.475, p=0.001	rs=-0.243, p=0.021	rs=-0.341, p=0.001
Abbreviations: SIP: Sickness Impact Profile questionnaire**; SF-36:** The 36-Item Short Form Survey**.**			

Table 7: Spearman's Correlation between SIP and SF36 at re-assessment (N=90).

Participant's self-assessment of health status

Participants were asked to grade their overall health status, post 1-week on a numeric rating scale (0-100%), where 0% indicated no degree of improvement and 100% indicated a very high degree of improvement. The percentage change (Post-Pre value/Pre Value × 100%) from baseline to endpoint for SIP and SF-36 questionnaires were calculated in 1-week interval. There were significant correlations between subjective health improvement and change in SIP and SF-36 scores as shown in (Table 8). The negative correlations between SIP and health improvement is because SIP post values were smaller that pre values, as lower SIP scores mean lower impact of disease (greater improvement).

	Subjective Health improvement
SIP Physical component	rs= -0.684, p=0.001
SIP Psychosocial component	rs= -0.658, p=0.001
SIP Total score	rs= -0.808, p=0.001
SF36 Physical health	rs= 0.632, p=0.001
SF36 Mental health	rs= 0.608, p=0.001
SF36 Total score	rs= 0.679, p=0.001
Abbreviations: SIP: Sickness Impact Profile questionnaire**; SF-36:** The 36-Item Short Form Survey**.**	

Table 8: Spearman's correlation between subjective health improvement and change in SIP/SF36 scores.

Known group validity

Significant differences between groups were found for domain and total scores of both SIP and SF-36. In general, impact of disease was higher for cardiopulmonary patients in both SIP Physical and Psychological component as well as total score. Impact on musculoskeletal patients was higher compared to obese except for Physical component at initial assessment and Total score at re-assessment (Table 9).

The results for SF-36 were comparable to SIP. In general, obese patients showed higher functional capacity than the other two groups. Cardiopulmonary patients were significantly different from musculoskeletal patients only in terms of Physical health (Table 10).

		Initial Assessment		*Re-assessment*	
		Median (IQR)	**Significant differences**	**Median (IQR)**	**Significant differences**
SIP Physical component	**CARDIOPULMONARY**	22.99 (17.38)	Cardiopulmonary significantly different from the other two groups	18.94 (23.76)	All groups different from each other
	MUSCULOSKELETAL	14.88 (19.41)		5.97 (6.74)	
	OBESE	10.02 (5.81)		9.72 (5.67)	
SIP Psychosocial component	**CARDIOPULMONARY**	15.64 (7.66)	All groups different from each other	7.41 (5.73)	All groups different from each other
	MUSCULOSKELETAL	8.42 (8.15)		5.28 (4.21)	
	OBESE	0 (2.23)		0 (2.82)	
SIP Total score	**CARDIOPULMONARY**	23.6 (12.94)	All groups different from each other	15.88 (15.25)	Cardiopulmonary significantly different from the other two groups
	MUSCULOSKELETAL	14.59 (8.2)		6.32 (5.45)	
	OBESE	7.04 (3.03)		6.98 (3.33)	
SF-36 Physical Health	**CARDIOPULMONARY**	28.4 (22.35)	All groups different from each other	35 (15.9)	All groups different from each other
	MUSCULOSKELETAL	33.6 (11.65)		46.6 (13.75)	
	OBESE	52.6 (13)		52.1 (18.25)	
SF-36 Mental Health	**CARDIOPULMONARY**	49.62 (13.07)	Obese significantly different from the other two groups	62.1 (17.17)	Obese significantly different from the other two groups
	MUSCULOSKELETAL	47.85 (8.37)		54.45 (17.14)	
	OBESE	75.4 (9.17)		74.9 (10.1)	
SF-36 Total score	**CARDIOPULMONARY**	35.17 (16.19)	Obese significantly different from the other two groups	48.17 (15.42)	Obese significantly different from the other two groups
	MUSCULOSKELETAL	36.13 (10.28)		47.72 (16.66)	
	OBESE	64.84 (11.98)		64.16 (13)	

Table 9: Comparisons between patient groups at initial assessment and re-assessment.

Discussion

The aim of this study was to assess the validity and reliability of the Greek version of SIP questionnaire in individuals with obesity, cardiac, pulmonary and musculoskeletal problems. The translation process followed published guidelines and was smooth without major problems. Overall, the results demonstrate that SIP-GR has acceptable validity and reliability, supporting the use of this questionnaire in the evaluation of QOL of patients with various problems.

Internal consistency was high for both Physical and Psychosocial domain as well as the total score. The German SIP version [18] reported slightly lower Cronbach's alpha for SIP total score (a=0.83) in patients with musculoskeletal disorders. Even lower value was reported in a study assessing polytrauma patients with lower extremity injuries (a= > 0.70) [19], whereas similar internal consistency (a=0.93) to this study, was reported by Hutter and Wurtemberger [20] in patients with COPD. Similarly, high value was found in the original version by Bergner et al. [9] (a=0.94) [20] and a study (a= >0.8) examining patients with Huntington's disease [21]. Recently, Majstorovic et al. assessed the reliability of the Serbian version of SIP in patients with chronic viral hepatitis and stated that the Cronbach's alpha for total score was 0.92, 0.86 for the physical dimension and 0.85 for the psychological dimension [22]. The Greek version demonstrated similar values for the overall score and comparable values for the domain scores. In addition, the Chinese version of SIP reported that the overall internal consistency of the questionnaire was 0.98, with no values below 0.70 for the category scores [5]. Moreover, high Cronbach's α from the SIP-total score [3], physical and psychological dimension scores are reported in other studies [21]. Table 10 shows a comparison of internal validity among different translations of SIP.

In addition, the SIP-GR demonstrated overall moderate to substantial test re-test variability. Test–retest results from studies reported an ICC= 0.94 for the physical function dimension and ICC= 0.93 for the overall SIP-136 score in patients with musculoskeletal disorders who completed a second SIP 3-weeks after the initial administration [23]. Ho et al. [21] reported reliability of the SIP scales (ICC= 0.70) in patients with Huntington disease, similar to this study (Table 9). The Italian version demonstrated that the majority of the test-retest correlations fell almost always within the 0.70-0.90 range [24]. Another study stated that the ICC for SIP-total score was 0.70 over a 2-week test–retest assessment in patients with chronic low back pain [25], compared to the 0.999 ICC of 7-days interval of this study for those that reported no change in health status. Table 10 shows a comparison of test-retest reliability among different translations of SIP. A relatively long interval between the two assessments was not considered as methodologically appropriate in this study, especially because most patients were receiving treatment. The effect of the treatment is evident by the change in scores in both SIP and SF-36. As expected, reliability was excellent in those patients that reported

no subjective change in their health status but moderate overall as most patients reported a change in their health status. The instrument seems to be sensitive to change as there was significant correlation between subjective health improvement and change in SIP scores (Table 8). In addition, only the patients with subjective health improvement showed significant differences between the initial and re-assessment values of both questionnaires (Table 5).

Version	*Internal consistency*			*Test-retest*		
	SIP-Overall	**Physical dimension**	**Psychological dimension**	**SIP-Overall**	**Physical dimension**	**Psychological dimension**
Japanese	0.96	0.95	0.93	0.92	0.91	0.90
American	0.94	-	-	0.92	-	-
Chinese	0.98	-	-	0.75*	-	-
Spanish	0.93-0.95‖	0.89-0.93‖	0.89-0.93‖	0.88†	-	-
Serbian	0.93	0.87	0.86	-	-	-
Swedish	0.95	-	-	0.91	0.91	0.87
Dutch	0.91	0.90	0.84	-	-	-
German	0.83	-	-	0.81	-	-

*Mean reliability of the category scores, ‖ between different ethnic groups, † in a small cohort of 10 patients only, Data for Swedish and Dutch versions were taken from [3].

Table 10: Different versions of SIP.

SIP-GR scores showed a significant negative correlation to SF-36 scores (Tables 6 and 7). Higher negative correlations were found between Physical component of SIP and Physical Health of SF-36 compared to Mental Health. Similarly, Psychosocial component of SIP showed higher correlation with Mental Health of SF-36 than the Physical Health. Correlations between domain and total scores of SIP and SF-36 were higher at initial assessment compared to follow up. This is difficult to explain as the pre and post comparisons of both questionnaires showed the same significant differences and there was correlation between subjective health improvement and changes in both SIP and SF-36. Considering the correlations of subjective health improvement and change in SIP and SF-36 total scores it seems that SIP demonstrate a higher correlation with subjective change. Perhaps treatment affected the scores of SIP more that the scores of SF-36 and this lowers the correlation of the two questionnaires at re-assessment.

Previous studies assessed the validity of SIP questionnaire using different disease specific instruments, such as: Rolland-Morris scale [26], Psoriasis Disability Index [27], Keitel Index [18], Oxford-12 [28]. No disease specific scales were used in this study as the intention was to assess the properties of SIP as a generic instrument. Therefore, a generic instrument such as SF-36 was used as a reference standard. In addition, SIP-GR demonstrated a moderate overall correlation to SF-36 questionnaire during re-assessment. One study utilized both SIP and SD-36 in polytrauma patients but the correlations are not included in the published report [19]. Another study although did not report the correlation between the two instruments reported a high similarity of carer's responses on QoL dimensions on both questionnaires [21].

Minimum detectable change is critical when judging the benefit of an intervention as it shows the level of change above measurement error that can be detected therefore is likely to be true treatment effect [29]. It also helps in the interpretation of the size of the treatment effect. In the physical domain, MDC was reported as 5 points, whereas in the psychological domain was 8 to 11 points in patients with COPD [30-32]. Moreover, it is reported that SIP-136 total score had high specificity to detect a change of 3 points in patients with rheumatoid arthritis [33,34]. Current study showed MDC of 4.6-5.5 points for the overall score, 6.4-7 for the Physical component and 3.5-4.9 for the Psychosocial component, at initial assessment and re-assessment respectively. The differences in MDC between this study and the previous studies can be explained by the methodological differences of the various studies.

Limitations

The main limitation of this study is the sample size. Although the power was sufficient for Test rest reliability, was not sufficient to perform an exploratory factor analysis to assess the components of the Greek version of SIP. The other limitation is that most patients were receiving treatment and they demonstrated a subjective health improvement because of it. However, this was taken into account and there was an effort to partial the effect of treatment wherever this was possible. Last but not least, no disease

specific scales were used to investigate the validity of SIP-GR. This was due to the aim of the study to investigate the validity of SIP-GR as a generic HRQOL tool. Future studies should address that in different populations.

Conclusion

This study describes the cultural adaptation, reliability and validity of the SIP-GR. The analysed psychometric properties showed that SIP-GR was reliable, valid and sensitive to change, and can therefore be recommended for clinical purposes. Further studies with larger samples are needed to validate the findings.

References

1. Lambert CM, Hurst NP (1995) Health economics as an aspect of health outcome: Basic principles and application in rheumatoid arthritis. Br J Rheumatol 34: 774-780.
2. Lillegraven S, Kvien TK (2007) Measuring disability and quality of life in established rheumatoid arthritis. Best Pract Res Clin Rheumatol 21: 827-840.
3. de Bruin AF, de Witte LP, Stevens F, Diederiks JP (1992) Sickness Impact Profile: The state of the art of a generic functional status measure. Soc Sci Med 35: 1003-1014.
4. Patrick DL, Deyo RA (1989) Generic and disease-specific measures in assessing health status and quality of life. Med Care 27: S217-232.
5. Short TG, Rowbottom MY, Lau JP, Lai GW, Buckley TA, Oh TE (1998) Translation and calibration of a Chinese version of the Sickness Impact Profile for use in Hong Kong. Hong Kong Med J 4: 375-381.
6. Pollard B, Johnston M (2001) Problems with the Sickness Impact Profile: A theoretically based analysis and a proposal for a new method of implementation and scoring. Soc Sci Med 52: 921-934.
7. Bujang MA, Baharum N (2017) A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. J Archives of Orofacial Science 12: 1-11.
8. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 25: 3186-3191.
9. Bergner M, Bobbitt RA, Carter WB, Gilson BS (1981) The Sickness Impact Profile: Development and final revision of a health status measure. Med Care 19: 787-805.
10. Sullivan M, Ahlmen M, Bjelle A (1990) Health status assessment in rheumatoid arthritis. I. Further work on the validity of the Sickness Impact Profile. J Rheumatol, 17: 439-447.
11. Matcham F, Norton S, Steer S, Hotopf M (2016) Usefulness of the SF-36 Health Survey in screening for depressive and anxiety disorders in rheumatoid arthritis. BMC Musculoskelet Disord 17: 224.
12. Taft C, Karlsson J, Sullivan M (2001) Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res 10: 395-404.
13. Bunevicius A (2017) Reliability and validity of the SF-36 Health Survey Questionnaire in patients with brain tumors: A cross-sectional study. Health Qual Life Outcomes 15: 92.
14. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60: 34-42.
15. Martinez BR, Lopes Sauers AD, Ferreira CL, de Castro Lugli L, Gama Turchetto PC, et al. (2018) Translation, cross-cultural adaptation, and measurement properties of the Brazilian version of the Identification of Functional Ankle Instability (IdFAI) questionnaire. Phys Ther Sport 29: 1-8.
16. Ries JD, Echternach JL, Nof L, Gagnon Blodgett M (2009) Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with Alzheimer disease. Phys Ther 89: 569-579.
17. Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M (2007) What we need to know when calculating the coefficient of correlation? Biochemia Medica 17: 10-15.
18. Kessler S, Jaeckel W, Cziske R (1997) Assessing health in musculoskeletal disorders: The appropriateness of a German version of the Sickness Impact Profile. Rheumatol Int 17: 119-125.
19. Jansen L, Steultjens MP, Holtslag HR, KwakkelG, Dekker J (2010) Psychometric properties of questionnaires evaluating health-related quality of life and functional status in polytrauma patients with lower extremity injury. J Trauma Manag Outcomes 4: 7.
20. Hutter BO, Wurtemberger G (1997) Quality of life in patients with chronic obstructive lung diseases: Psychometric criteria of the German version of the Sickness Impact Profile and initial results of its application. Pneumologie 51: 108-114.
21. Ho AK, Robbins AO, Walters SJ, Kaptoge S, Sahakian BJ, Barker RA (2004) Health-related quality of life in Huntington's disease: A comparison of two generic instruments, SF-36 and SIP. Mov Disord 19: 1341-1348.
22. Majstorovic B, Jankovic S, Dimoski Z, Kekus D, Kocic S, Mijailovic Z (2015) Assessment of the reliability of the serbian version of the Sickness Impact Profile questionnaire in patients with chronic viral hepatitis. Srp Arh Celok Lek 143: 688-694.
23. Beaton DE, Hogg-Johnson S, Bombardier C (1997) Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 50: 79-93.
24. Bertolotti G, Vidotto G, Baiardi P, Carone M, Sommaruga M, Zotti AM (2001) Sickness Impact Profile: The Italian version. G Ital Med Lav Ergon 23: 477-483.
25. Deyo RA, Diehr P, Patrick DL (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 12: 142S-158S.
26. Jensen MP, Strom SE, Turner JA, Romano JM (1992) Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. Pain 50: 157-162.
27. Finlay AY, Khan GK, Luscombe DK, Salek MS (1990) Validation of Sickness Impact Profile and Psoriasis Disability Index in Psoriasis. Br J Dermatol 123: 751-756.
28. Dunbar MJ, Robertsson O, Ryd L, Lidgren L (2000) Translation and validation of the Oxford-12 item knee score for use in Sweden. Acta Orthop Scand 71: 268-274.
29. Stipancic KL, Yunusova Y, Berry JD, Green JR (2018) Minimally detectable change and minimal clinically important difference of a decline in sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis. J Speech Lang Hear Res 61: 2757-2771.
30. Bowers S, Cannizzaro K, Gregus J, Scott Q, Eason J (2009) Outcomes in cardiopulmonary physical therapy: Sickness Impact Profile. Cardiopulm Phys Ther J 20: 19-24.

31. MacKenzie CR, Charlson ME, DiGioia D, Kelley K (1986) Can the Sickness Impact Profile measure change? An example of scale assessment. J Chronic Dis 39: 429-438.

32. Stratford P, Solomon P, Binkley J, Finch E, Gill C (1993) Sensitivity of Sickness Impact Profile items to measure change over time in a low-back pain patient group. Spine (Phila Pa 1976) 18: 1723-1727.

33. Deyo RA, Inui TS (1984) Toward clinical applications of health status measures: Sensitivity of scales to clinically important changes. Health Serv Res 19: 275-289.

34. McKenna SP (2011) Measuring patient-reported outcomes: Moving beyond misplaced common sense to hard science. BMC Med 9: 86.