VALUE IN HEALTH 16 (2013) 1023–1031

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Exploring the Consistency of the SF-6D

Lara N. Ferreira, PhD[1,2,*], Pedro L. Ferreira, PhD[2,3], Luis N. Pereira, PhD[1,4], Donna Rowen, PhD[5], John E. Brazier, PhD[5]

[1]University of the Algarve-ESGHT, Faro, Portugal; [2]Centre for Health Studies & Research, University of Coimbra, Coimbra, Portugal; [3]Faculty of Economics, University of Coimbra, Coimbra, Portugal; [4]Research Centre for Spatial and Organizational Dynamics, University of the Algarve, Faro, Portugal; [5]Health Economics and Decision Science Section, School of Health and Related Research, University of Sheffield, Sheffield, UK

ABSTRACT

Objective: The six-dimensional health state short form (SF-6D) was designed to be derived from the short-form 36 health survey (SF-36). The purpose of this research was to compare the SF-6D index values generated from the SF-36 ($SF\text{-}6D_{SF\text{-}36}$) with those obtained from the SF-6D administered as an independent instrument ($SF\text{-}6D_{Ind}$). The goal was to assess the consistency of respondents' answers to these two methods of deriving the SF-6D. **Methods:** Data were obtained from a sample of the Portuguese population (n = 414). Agreement between the instruments was assessed on the basis of a descriptive system and their indexes. The analysis of the descriptive system was performed by using a global consistency index and an identically classified index. Agreement was also explored by using correlation coefficients. Parametric tests were used to identify differences between the indexes. Regression models were estimated to understand the relationship between them. **Results:** The $SF\text{-}6D_{Ind}$ generates higher values than does the $SF\text{-}6D_{SF\text{-}36}$. There were significant differences between the indexes across sociodemographic groups. There was a significant ceiling effect in the $SF\text{-}6D_{Ind}$ but not in the $SF\text{-}6D_{SF\text{-}36}$. The correlation between the indexes was high but less than what was anticipated. The global consistency index identified the dimensions with larger differences. Considerable differences were found in two dimensions, possibly as a result of different item contexts. Further research is needed to fully understand the role of the different layouts and the length of the questionnaires in the respondents' answers. **Conclusions:** The results show that as the SF-6D was designed to derive utilities from the SF-36 it should be used in this way and not as an independent instrument.

Keywords: consistency, dimensions, SF-6D, SF-36.

Introduction

The short-form 36 health survey (SF-36) is a 36-item generic health status instrument, comprising eight scales and two component summary scales [1,2], which has been extensively validated and used e.g., [3–12]. The SF-36 is a profile-based patient-reported outcome measure that yields health scores across its eight dimensions. It does not, however, generate utilities, and hence it has a limited use in economic evaluations of health care interventions or technologies. To overcome this problem, a decade ago, Brazier et al. [13] developed an algorithm to translate the SF-36 results into health state utilities. They created the six-dimensional health state short form (SF-6D), an econometric preference-based index derived from 11 items of the SF-36, which are combined into six dimensions of health, with four to six levels each [13]. The SF-6D describes 18,000 different health states. A valuation survey was carried out in the United Kingdom to obtain values to a sample of 249 health states defined for the SF-6D. A representative sample of the general UK population valued these health states by using the standard gamble method. Econometric models were estimated by using the data collected to predict utility scores for all health states defined by the SF-6D [14]. These health state values constitute the SF-6D index, which can be seen as a continuous value ranging from 0.35 to 1.00. Another version of the SF-6D was developed on the basis of the short-form 12 health survey instrument (SF-12), and utility scores for all health states defined by this instrument for the UK population are also available [14]. The SF-6D enables a utility score to be generated by using responses to the SF-36 or the SF-12. There are now specific value sets for the SF-6D for Portugal [15], Japan [16], Hong Kong [17], and Brazil [18], with value sets for Australia and Singapore currently being determined.

Given that the SF-36 is widely used all over the world, the use of the SF-6D as a way of generating utilities from the SF-36 has increased in recent years and is now one of the preference-based indexes most widely used in cost-utility analyses and other studies that aim at measuring individuals' preferences for health states [e.g., 19–26] and included in numerous pharmacoeconomics guidelines. Previous research has focused on the assessment of the performance of the SF-6D and on comparisons with other

* *Address correspondence to*: Lara N. Ferreira, Escola Superior de Gestão, Hotelaria e Turismo, Universidade do Algarve, Campus da Penha, 8005-139 Faro, Portugal.

E-mail: Lnferrei@ualg.pt.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.06.018

preference-based indexes such as the EuroQol five-dimensional questionnaire or the Health Utilities Index. Several articles have been published on these topics [e.g., 27–34]. It is also useful to know whether the SF-6D health state classification can be used in its own right because this would be a more efficient way to collect the data. There are, however, no published studies having been dedicated to studying the consistency of the SF-6D when the classification system is used directly in a study to derive the health state of the individual rather than deriving the health state from responses to the SF-36 or the SF-12. Therefore, we intend to overcome this gap in the literature by exploring the consistency of respondents' answers to these two methods of deriving the SF-6D.

The aim of this research was twofold: 1) to test the hypothesis that the SF-6D applied as an independent instrument ($SF\text{-}6D_{Ind}$) produces results different from those obtained from the SF-6D index generated from the SF-36 ($SF\text{-}6D_{SF\text{-}36}$) and 2) to examine whether the conclusions differed depending on the value set used.

Methods

Sample and Data Collection

Data were collected from a sample of individuals from the adult general Portuguese (PT) population (n = 414) in spring 2011 in Portugal. Respondents were recruited from the population of students and staff of a public university in Portugal, according to their willingness to participate in the study. Although the sample used in the study is nonrandom, it was expected to include respondents from different sociodemographic groups given that the population of individuals comprised undergraduate and graduate students and teaching and nonteaching staff. In addition, it was not essential to use a random representative sample of the general population given that to achieve the aim of this research it was only necessary to prove that using the $SF\text{-}6D_{Ind}$ produces results different from those obtained from the $SF\text{-}6D_{SF\text{-}36}$ in at least one sample.

Respondents self-completed the SF-36v2 and the SF-6D on a voluntary and anonymous basis. This enables analysis of the consistency of respondents' answers to the two above-mentioned methods of deriving the SF-6D index; the SF-6D was also applied as an independent questionnaire.

The order of the self-completed paper-and-pencil questionnaires was fixed and was the same throughout the study: first, the SF-36, and second, the SF-6D classification system. In addition, respondents reported information on sociodemographic variables, such as sex, age, marital status, education, labor market participation, area of living, income, and the presence (or not) of a chronic disease.

The UK [14] and the PT [15] value sets for the SF-6D were both applied to the data collected to further examine whether the conclusions differed depending on the value set used. We have applied only these two value sets because there are no other European value set for the SF-6D and the UK value set is considered the gold standard. In fact, before the elicitation of the PT value set, studies conducted in Portugal used UK population values.

Statistical Analysis

Sample characteristics were first described by computing descriptive statistics for sociodemographic variables. The analysis of the degree of agreement between instruments was divided into two parts. First, an analysis based on the classification system of both instruments was performed, that is, an analysis of what respondents reported about their health in each instrument. This task started with a general descriptive analysis of the distribution of responses across dimensions in both instruments. Then, the degree of association between dimensions of the SF-6D was measured by using the Spearman's correlation coefficient. In addition, we used the following two measures based on square two-way contingency tables: a global consistency index (GCI) and an identically classified index (ICI). The GCI computes the percentage of individuals classified in the same level of each dimension in both instruments and is given by

$$GCI = \frac{\sum_{j=1}^{l} n_{jj}}{n} \times 100, \quad (1)$$

where n is the sample dimension and n_{jj} is the number of individuals with response in the same level j (j = 1, …, l) of a particular dimension in the $SF\text{-}6D_{SF\text{-}36}$ and in the $SF\text{-}6D_{Ind}$. The GCI will be equal to 100 if all individuals equally respond on a specific dimension in both instruments. GCI values above 75 are interpreted as a strong agreement between instruments, whereas GCI values ranging from 50 to 75 are considered as a moderate agreement. GCI values lower than 50 suggest a poor agreement. The ICI calculates the percentage of individuals correctly classified in a level j of each dimension in the $SF\text{-}6D_{Ind}$ and is given by

$$ICI_j = \frac{n_{jj}}{n_{j\bullet}} \times 100, \quad (2)$$

where $n_{j\bullet} = \sum_{k=1}^{l} n_{jk}$ is the total number of responses in level j of a particular dimension in the $SF\text{-}6D_{SF\text{-}36}$. The ICI can be interpreted as a stability indicator and will be equal to 100 if all individuals equally respond on a level j of a specific dimension in both instruments. We also define a poor level of stability on responses when the ICI is less than or equal to 25.

Second, an analysis of the preference-based indexes generated by the instruments was carried out by using the following data analysis: 1) basic descriptive statistics including means, medians, and ranges to compare the main features of the indexes; 2) skewness statistics and one-sample Kolmogorov-Smirnov tests to evaluate the asymmetry and normality of distributions; 3) ceiling and floor effects (proportion of respondents with the best and worst possible theoretical scores, respectively) were identified; 4) Pearson's correlation coefficients to study the association between instruments and intraclass correlation coefficients (ICCs) based on a two-way mixed model with absolute agreement, for a global assessment of the agreement between indexes; 5) paired-samples t test (related-samples Wilcoxon signed-rank tests) to identify mean (median) differences between the indexes; and 6) regression analysis to explore the nature of their relationship. We have used the following model:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad (3)$$

where Y represents the $SF\text{-}6D_{SF\text{-}36}$ index, X the $SF\text{-}6D_{Ind}$ index, ε the residuals, and i respondents (i = 1, …, n). It should be noted, however, that the aim of the regression analysis was to test whether there is a perfect agreement between the indexes and not to explain or predict the $SF\text{-}6D_{SF\text{-}36}$ index through the $SF\text{-}6D_{Ind}$ index. Because an agreement between the indexes would result in estimated models in which the constant (α) would be equal to zero and the slope (β) equal to one, hypothesis tests were performed to verify these assumptions. Finally, the pattern of agreement was also examined graphically by plotting values obtained for the UK and PT value sets.

It should be noted that mean differences between indexes were evaluated by using paired-samples t tests, although the normality assumption was not verified. This decision was based on the following: the large sample size of our study; the well-known result that the power of the Kolmogorov-Smirnov Z test increases with the sample size; and some evidence that