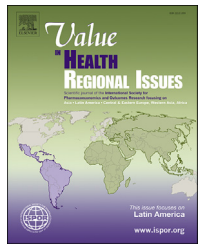
Available online at www.sciencedirect.com

ScienceDirect

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



Preference-Based Assessments

A Comparison of UK and Brazilian SF-6D Preference Weights When Applied to a Brazilian Urban Population

Marcos Bosi Ferraz, MSc, PhD [1], *Elene Paltrinieri Nardi, MSc* [1,*], *Alessandro Gonçalves Campolina, MSc, PhD* [2]

[1]*Disciplina de Economia e Gestão em Saúde, Universidade Federal de São Paulo, São Paulo, Brazil;* [2]*Centro de Investigação Translacional em Oncologia, Instituto do Câncer do Estado de São Paulo, Universidade de São Paulo, São Paulo, Brazil*

ABSTRACT

Background: The 6-dimensional health state short form (SF-6D) is a health preference measure used in economic evaluations of many treatments. **Objectives:** To compare the results provided by the SF-6D index, when applied to a representative sample of the Brazilian population, using Brazilian and UK preference weights. **Methods:** Five thousand individuals were assessed in the 5 regions of Brazil. Preference measures in healthcare were assessed using the SF-6D Brazil, version 2002. To calculate the single utility score, 2 preference weights were used: one established for the Brazilian population (SF-6D Brazil) and the other for the UK population (SF-6D UK). Agreement between the SF-6D Brazil and the SF-6D UK was assessed using the intraclass correlation coefficient, the Wilcoxon signed rank test, confidence intervals (CIs), and the Bland-Altman method. **Results:** The mean values of the SF-6D Brazil and the SF-6D UK were 0.83 ± 0.15 and 0.84 ± 0.15, respectively. The intraclass correlation coefficient was 0.952 (CI 0.942-0.960; $P<.010$). The Wilcoxon signed rank test and CI showed a statistically significant difference between the 2 measures; this difference was, however, very small and considered clinically irrelevant (CI 0.011-0.013; $P<.010$). Using the Bland-Altman method resulted in a mean difference of 0.012 and the limits of agreement were between −0.077 and 0.101. **Conclusions:** The present study identified very small quantitative differences between UK- and Brazilian-derived SF-6D scores. Tests of agreement, however, showed that the impact of using different sets of preference weights in the construction of quality-adjusted life-year might be considered irrelevant.

Keywords: Brazil, health status, quality-adjusted life-years, quality of life, SF-6D

Introduction

One of the primary goals of the healthcare system is to maximize health benefits given that resources are scarce. Economic evaluations can provide vital information on how to efficiently allocate these resources.[1]

Cost-effectiveness analysis using cost per quality-adjusted life-year (QALY), also known as cost-utility analysis, is increasingly being used to inform resource allocation in healthcare. QALYs are defined as the benefit of a health intervention in terms of time in a series of quality-weighted health states, in which the quality weights reflect the desirability (or preference) of living in the particular health state, typically anchored at "perfect health" (weighted 1.0) and "dead" (weighted 0.0).[2]

The use of QALY is recommended in economic evaluation analyses of medical interventions by both the National Institute for Health and Care Excellence in the United Kingdom[3] and the Panel on Cost-Effectiveness in Health and Medicine of the US Public Health Service,[4] because it enables comparisons across a wide range of diseases and treatments using a common measurement.

The "quality" part of the QALY is estimated using a preference-based measure of health, such as the EuroQol 5-dimensional questionnaire (EQ-5D), the Health Utilities Index, and the 6-dimensional short form health survey (SF-6D).[5] These are all instruments to provide utilities (values representing preferences) for healthcare priority settings typically obtained indirectly by asking patients to fill in a health-related quality of life (HRQOL) questionnaire and then converting the results to a utility scale using population values.[5]

The SF-6D is a popular health preference measure that is used in economic evaluations of many treatments. It is a 6-dimensional health state classification system that describes a total of 18 000

Conflicts of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Elene Paltrinieri Nardi, MSc, Disciplina de Economia da Saúde, Universidade Federal de São Paulo, Rua Botucatu 740, São Paulo CEP-04023-062, Brazil.

Email: epnardi@gmail.com

2212-1099/$36.00 - see front matter © 2018 ISPOR–The professional society for health economics and outcomes research. Published by Elsevier Inc.

https://doi.org/10.1016/j.vhri.2018.10.009

distinct health states. Dimensions include physical functioning, role limitations, social functioning, pain, mental health, and vitality. Each dimension has between 4 and 6 levels. The SF-6D produces a preference-based single index utility value for each health state anchored on the full health to death 1 to 0 scale, which can be seen as a continuous value ranging from 0.29 to 1.00 (UK value set).[6,7]

Each instrument of this kind consists of a multiattribute classification system for distinguishing health states and a scoring function that assigns a value to each health state distinguished within the classification system.[8] In valuation studies, the method to estimate a value set, to translate health outcomes into a preference-based measure, involves the following steps: (1) systematic description of health states, by dimensions and levels; (2) selection of a subset of health states from all the possible health states; (3) quantification of public preferences regarding the subset states; and (4) modeling the obtained preference data to predict the preference regarding the remaining health states.[9] The size of the identified benefit in economic evaluations depends on the normative choices of the patients and the general public regarding their preferences for health states.[10]

This approach is sometimes referred to as the statistical approach and has been applied to determine scoring functions for the SF-6D.[8] It assumes that the social value or "utility" of a health state is the same as the value of the quality of life of those individuals/general public who are in it, and that the description of the health state of a subject may be summarized, in theory at least, as a health index that reflects the social preferences of the population. Nevertheless, the crucial question in seeking social preference weights for items is how much difference it makes to use these differential weights to calculate the composite utility score for a given health state.

In a previous valuation study, the association between weighted and unweighted EQ-5D scores was extremely high (Pearson $r = 0.91$), as was the association between their ranks (Spearman $\rho = 0.93$). The intraclass correlation coefficient (ICC) obtained (0.89) also suggested that the concordance between the score distributions was prominent. According to the authors, the differential contribution of weights based on population preference values would therefore be minimal between countries.[11]

Nevertheless, there has been increasing interest in developing country-specific preference weights for widely used measures of HRQOL.[12–16] Studies evaluating the EQ-5D and the SF-6D in diverse cultures have shown that health state preferences are different from the preferences derived in the country where the measure was originally created.[9,13,14,17]

The objective of this study was to compare the generated utility values provided by the SF-6D index, when applied to a representative sample of the Brazilian population, using the Brazilian and the UK preference weights.

Methods

This study was a cross-sectional population-based study. It was part of the Brazilian Copcord Study, conducted between April and May 2013. We surveyed 5000 participants who were older than 15 years from 16 capitals in the 5 regions of Brazil: North (Belém and Manaus), Northeast (Fortaleza, João Pessoa, Maceió, Natal, Recife, and Salvador), Southeast (Belo Horizonte, Rio de Janeiro, and São Paulo), South (Curitiba, Florianópolis, and Porto Alegre), and Midwest (Brasília and Goiânia).

The sample comprised representative quotas of the Brazilian population, proportional to the population densities of the capitals in each region of the country, on the basis of the Census conducted in 2010 by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*). The quotas of sex and age in each capital were based on the Census, and participants of all socioeconomic status, educational levels, and occupations were included.

The households were randomly selected, with a systematic selection of streets and subjects by randomly choosing the Census tract with a quota control for the seasonality factor. Regarding the list of households, 1 household was evaluated per street, with a total of up to 10 households in the sector. If an entire sector was covered but not enough households were found to complete the required number, the process was carried out again in the sector, beginning in the first street, 5 houses after the house where the first interview took place. In each household, up to 3 visits were made on different days and at different times. In cases in which the interview was not carried out after these 3 attempts, the household was replaced by another in the same Census sector. If the resident of the selected household could not be interviewed, that household was replaced by another in the same Census sector, seeking to ensure a respondent of the same sex and age group.

A success rate of 70% was established, so 42.9% more households than planned were randomly selected to ensure substitution. Ineligible households, such as collective households (vacant households, hotels, lodges, nursing homes, etc), agricultural, educational, and healthcare establishments, and buildings under construction, were replaced by another household in the same Census sector. The maximum sampling error was ±1.39% for the country as a whole, with a 95% confidence level.

Residents who did not speak Portuguese and people with a cognitive disability thus incapable of reliably and consistently answering the questionnaire were excluded. Because of the small proportion of people living in rural areas (15.6%) and the difficulty in accessing this scattered population, only households in urban areas were considered.

The questionnaires were reviewed by an independent supervisor and submitted to a process to evaluate consistency, where 50% of the questionnaires were double-checked through phone calls.

The survey instrument was a household questionnaire administered face-to-face by a specialized team. It consisted of open- and closed-ended questions about socioeconomic and demographic aspects. Race (white, black, mixed race, Asian, or indigenous) was declared by the respondents themselves. Family income was expressed as multiples of minimum wages, where the values of the Brazilian minimum wage, originally in Brazilian real (Brazilian currency), were converted into US dollars according to the exchange rate in 2013 using data from the Institute of Applied Economic Research (*Instituto de Pesquisa Econômica Aplicada*).

Preference measures in healthcare were assessed using the SF-6D Brazil questionnaire from 2002, a translated and validated version for the Portuguese language.[18] The dimensions assessed were functional capacity (6 levels), general limitation (4 levels), social aspects (5 levels), pain (6 levels), mental health (5 levels), and vitality (5 levels). To calculate the single utility score, 2 preference weights were used: the weight established for the Brazilian population and described by Cruz et al[12] and the weight established for the UK population and described by Brazier et al.[6] The utility values measured using the Brazilian preference weight (SF-6D Brazil) range from 0.289 (worst health state) to 1 (best health state) and those measured using the UK preference weight (SF-6D UK) range from 0.296 to 1.[6,12]

The variables were descriptively analyzed, and data were presented as mean, SD, and percentage. Agreement between the SF-6D Brazil and the SF-6D UK was assessed using ICC, the Wilcoxon signed rank test, and confidence intervals (CIs) for the differences between means of utility scores. Both preference weights were considered to agree when ICC was more than 0.75 ($P>.050$; Wilcoxon test) and if the 95% CI contained 0.[19] The Bland-Altman method was used as statistical criteria to establish disagreement

Table 1 – Demographic and socioeconomic characteristics of the surveyed population (n = 5000) and the Brazilian population (Census 2010).

Characteristic	Brazilian population[40]		Participants	
	N	%	n	%
Sample total	190 755 799	100	5000	100
Sex				
Male	93 406 990	49.0	2433	48.7
Female	97 348 809	51.0	2567	51.3
Age* (y)				
15-24	34 236 064	17.9	1270	25.4
25-34	32 848 926	17.2	1160	23.2
35-44	26 897 943	14.1	915	18.3
45-54	21 973 754	11.5	692	13.8
55-64	14 785 341	7.8	490	9.8
≥65	14 081 477	7.4	473	9.5
Marital status				
Single	89 653 403	55.3	2053	41.1
Married/lives with partner	56 435 253	34.8	2442	48.8
Widowed	8 063 404	5.0	240	4.8
Divorced/separated	7 829 238	4.8	233	4.7
Not reported	–	–	32	0.6

* ≥10 y (n = 161 981 299).

between the SF-6D Brazil and the SF-6D UK.[20] This method quantifies the gap between the x-axis corresponding to zero difference and the x-axis corresponding to mean difference of the 2 SF-6D utility scores and range of agreement (the range of values where 95% of differences between the 2 SF-6D utility scores are included).[21]

Statistical analyses were performed using the STATISTICA software for Windows version 12.0 (StatSoft, Inc, Tulsa, OK, USA), and statistical significance was assumed for *P* values of .050 or less.

All subjects were informed about the study and agreed to participate by signing a written term of free and informed consent. The research protocol was examined and approved by the Ethics and Research Committee of UNIFESP/EPM (no. 2013/473524).

Results

Table 1 presents the main demographic and socioeconomic characteristics of the surveyed population. Of the 5000 participants, 51.3% were women, 32% married, and 48.6% between 15 and 34 years old. The sample distribution by region of the country was proportional to the population densities of the capitals in each region of the country, on the basis of the Census conducted in 2010. The highest prevalence of the surveyed population was in the Southeast (42.1%) (Table 1).

The utility values measured by the SF-6D in the study population are presented in Table 2. The mean values of the SF-6D Brazil and the SF-6D UK for the entire population were 0.83 ± 0.15 and 0.84 ± 0.15, respectively. Regarding characteristics of the surveyed population, the lowest utility values were for females, the elderly, and widows (Table 2).

Both preference weights were considered to agree, because the ICC was 0.952 (CI 0.942-0.960; P<.010) (Fig. 1). The Wilcoxon signed rank test and CI for the differences between means of utility scores showed a statistically significant difference between the 2 measures (CI 0.011-0.013; P<.010). Nevertheless, the mean difference was minimal and clinically irrelevant (Table 2). Regarding characteristics of the surveyed population, minimum differences were also obtained between the means across sex, age, and marital status (Table 2).

The Bland-Altman method demonstrated that the mean difference between the paired data was 0.012 (CI 0.011-0.013). This marginal CI does not include the x-axis corresponding to zero difference, which means that there was a significant difference between both preference weights, albeit very small (Fig. 2).

Table 3 presents individual SF-6D values for a sample of 30 health states of different levels of severity (out of 1357 possible health states).

Discussion

Discrepancies between utilities predicted by HRQOL measures, using different sets of preference weights, have been observed in other studies.[22–24] This is in agreement with previous research that suggests that systematic differences exist in health state valuations across countries when using the SF-6D index.[9,14,17,25,26]

The present study, however, identified only minimal quantitative differences, suggesting that mean SF-6D values for a sample of the Brazilian population may not be largely affected by the use of different preference weights. Our data highlight a possible agreement between SF-6D scores, because high intraclass correlation and "fair" Bland-Altman agreement might suggest a similarity between values produced by different sets of preference weights.

In Japan, 1 study has also replicated the methods developed for estimating a preference-based measure for health in the United Kingdom.[14] The results showed differences between the UK and the Japanese samples. The Japanese valuation of the worst state was significantly higher than that for the United Kingdom (0.29 vs 0.21). According to the authors, because of these differences it is important to use the Japanese valuation data set estimated using the nonparametric Bayesian technique presented in the article.

Comparisons between the Portuguese and UK preference weights suggested that the SF-6D values for a sample of the Portuguese working-age population are also affected by the use of sets of preference weights.[25] Furthermore, the results of this valuation study suggested higher similarity between preference weights derived using rank data than using standard gamble.[13]

To ensure that the SF-6D can be appropriately used in economic evaluations and outcome studies, researchers have also sought to demonstrate test-retest reliability. In different populations, ICC has ranged from 0.66 to 0.94.[17,27–30] In a Canadian population, health state values were very similar between the first and second assessments; the mean difference was 0.01 (0.02-0.04).[27] Likewise, in China, the difference between the mean scores of the groups was −0.026 (95% CI −0.069 to 0.017), which was not shown to be statistically significant by the paired *t* test (t = −1.174; P=.242).[17] In most of these studies, reliability for the SF-6D was high, allowing for group comparison.[17,27]

Similarly, in the present study, the difference in SF-6D utility scores for different age groups also showed small differences depending on the set of preference weights used, with differences in each age category ranging from 0.01 to 0.03 (Table 2). These findings suggest only small differences between scores, because the minimally important difference (MID) for the SF-6D ranged from 0.010 to 0.048 (weighted mean 0.033; 95% CI 0.029-0.037) in a previous review of 9 studies.[31] It is important, however, to consider that the MID is crucial when the instrument is used to examine within-patient changes, but this does not necessarily mean that this is significant when the instrument is used to discriminate between patients.[31]

Table 2 – Comparison of the results provided by the SF-6D index using the system weights for Brazil (SF-6D Brazil) and for the United Kingdom (SF-6D UK).

Characteristic	N	SF-6D Brazil			SF-6D UK			P value*	CI [$\mu_{SF\text{-}6D\ UK}$-$\mu_{SF\text{-}6D\ Brazil}$]†
		Mean ± SD	Minimum	Maximum	Mean ± SD	Minimum	Maximum		
Sample total	5000	0.83 ± 0.15	0.30	1.00	0.84 ± 0.15	0.30	1.00	<.010	0.011 to 0.013
Sex									
Male	2433	0.86 ± 0.14	0.30	1.00	0.88 ± 0.13	0.30	1.00	<.010	0.009 to 0.012
Female	2567	0.79 ± 0.16	0.30	1.00	0.81 ± 0.15	0.30	1.00	<.010	0.012 to 0.015
Age (y)									
15-24	1270	0.88 ± 0.12	0.49	1.00	0.89 ± 0.12	0.43	1.00	<.010	0.002 to 0.006
25-34	1160	0.84 ± 0.13	0.43	1.00	0.86 ± 0.13	0.46	1.00	<.010	0.007 to 0.012
35-44	915	0.82 ± 0.15	0.36	1.00	0.84 ± 0.15	0.37	1.00	<.010	0.009 to 0.015
45-54	692	0.79 ± 0.16	0.30	1.00	0.80 ± 0.16	0.30	1.00	<.010	0.012 to 0.019
55-64	490	0.76 ± 0.17	0.33	1.00	0.79 ± 0.16	0.32	1.00	<.010	0.021 to 0.029
≥65	473	0.75 ± 0.17	0.30	1.00	0.78 ± 0.17	0.30	1.00	<.010	0.017 to 0.025
Marital status									
Single	2053	0.86 ± 0.14	0.30	1.00	0.87 ± 0.14	0.30	1.00	<.010	0.005 to 0.009
Married/lives with partner	2442	0.81 ± 0.15	0.30	1.00	0.83 ± 0.15	0.30	1.00	<.010	0.013 to 0.017
Widowed	240	0.74 ± 0.17	0.30	1.00	0.77 ± 0.17	0.30	1.00	<.010	0.018 to 0.031
Divorced/separated	233	0.78 ± 0.16	0.36	1.00	0.79 ± 0.16	0.35	1.00	<.010	0.005 to 0.018
Not reported	32	0.86 ± 0.15	0.52	1.00	0.87 ± 0.14	0.51	1.00	.112	−0.003 to 0.028

CI indicates confidence interval; SF-6D, 6-dimensional health state short form.
* *P* values were obtained by the Wilcoxon signed rank test.
† CIs for the differences between means of utility scores.

First, the degree of agreement (equivalence) between UK- and Brazilian-derived SF-6D values was determined using ICC (0.992). Second, assessment of agreement was based on the Bland-Altman plot.[32] Despite the relatively large limits of agreement in the Bland-Altman plot (−0.077 to 0.101), it is still lower than that found by Slobogean et al[27] in a study of test-retest reliability using the SF-6D (−0.17 to 0.19). Moreover, the statistical difference found in the Bland-Altman plot may be considered clinically irrelevant and can be explained by the large sample used in this study. The results, therefore, showed that there was a high degree of agreement between both derived SF-6D scores.

Other research has suggested that health state utilities measured by the major multiattribute utility instruments differ.[33–35] The reasons for discrepancies between utilities predicted by different preference-based instruments may be studied in terms of (1) their implicit measurement scales, (2) the structure of their descriptive systems, and (3) "micro-utility effects," scale-adjusted differences attributable to their utility formula.[35,36]

In a recent study, comparing the EQ-5D, the SF-6D, the Health Utilities Index, the 15D measure of quality of life, and the Assessment of Quality of Life, 66% of the differences between the predicted utilities was attributable to the descriptive systems, 30.3% to scale effects, and 3.7% to micro-utility effects.[35] The small micro-utility effects found in this study suggest that differences in utilities attributed to national preferences should be considered with great caution, at least until new results can be attributed to unique preferences, because the effects could also be related to differences in the methodologies used to derive the utility formula.[35] This means that a descriptive system is enough to explain a high proportion of variance in scores obtained through the use of preference-based measures and that weighting answers to the items in the instruments does not imply a significant difference. Thus, the final scores of these measures are fundamentally a direct reflection of the answers provided by the individuals to the items in the questionnaire.[11]

Considering that the ability of the models to adequately predict health state values[6,25,37] is more important than the magnitude of the differences in preference weights generated by different valuation methods, the recommended UK consistent model is different from the Brazilian one (Table 4). Statistically, these are not comparable models, because the UK model was estimated by ordinary least squares using data at a mean level, including the interaction term.[6,12] The recommended Brazilian model, however, was the consistent random-effects model with constant forced to unity.[12]

It seems that these modeling considerations have central implications for scoring functions when utilities are determined from standard gamble, time trade-off, or rating scale values using the statistical approach. It would therefore be possible to argue that only additive models with a freely estimated additive parameter should be applied.[8] Furthermore, similarity was also attributed to methodological aspects of the valuation process in the Portuguese (preference weights derived using rank data) and Japanese (nonparametric Bayesian technique) studies previously cited.[14,25]

The impact of these different sets of preference weights on incremental QALYs and incremental cost-utility ratios appears to be of major importance.[38] Nevertheless, some consider that "a systematic difference in the probability of accepting the cost-utility of interventions as a result of the choice of utility instrument would seriously bias the comparability of the results of economic evaluations."[1]

In this sense, in a previous study, the choice of a preference-based utility instrument among a population of community-dwelling older women would be relevant when comparing the cost-utility of 2 doses of resistance training.[38] Nevertheless, the US and UK QALY predictions for SF-6D states when compared visually and formally (Lin coefficient of agreement was 0.941 and mean absolute difference was 0.043) using the SF-6D responses from the United States (N = 259 243) were considered equivalent. The scatterplot in the study illustrated the

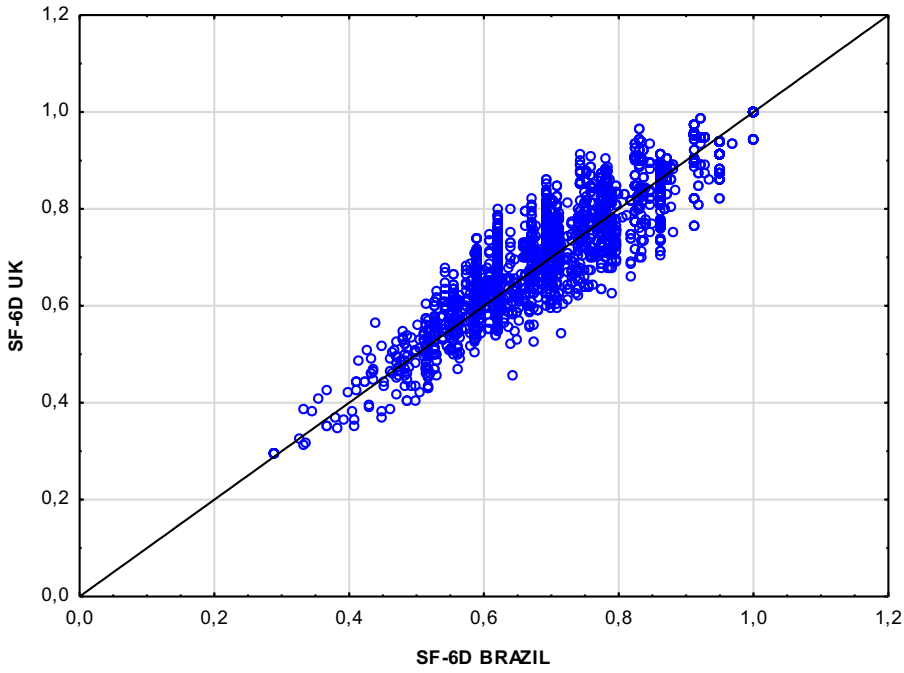


Fig. 1 – Scatter plot: SF-6D UK versus SF-6D Brazil.

commonality between the predictions even though the studies were conducted with different valuation tasks and sampling.[15]

The measurement of utilities is a complicated and controversial area,[5] and how these differences in the SF-6D scores could be attributed to preference weights is still unclear. Future research should (1) compare different SF-6D country-specific preference weighs (in different contexts); (2) take into account the small residual micro-utility effect, and explore its contribution to an explanation of differences, considering the variation in modeling and other methodological techniques; (3) evaluate the impact of different sets of SF-6D preference weights in the construct validity and responsiveness of the QALY metric, when

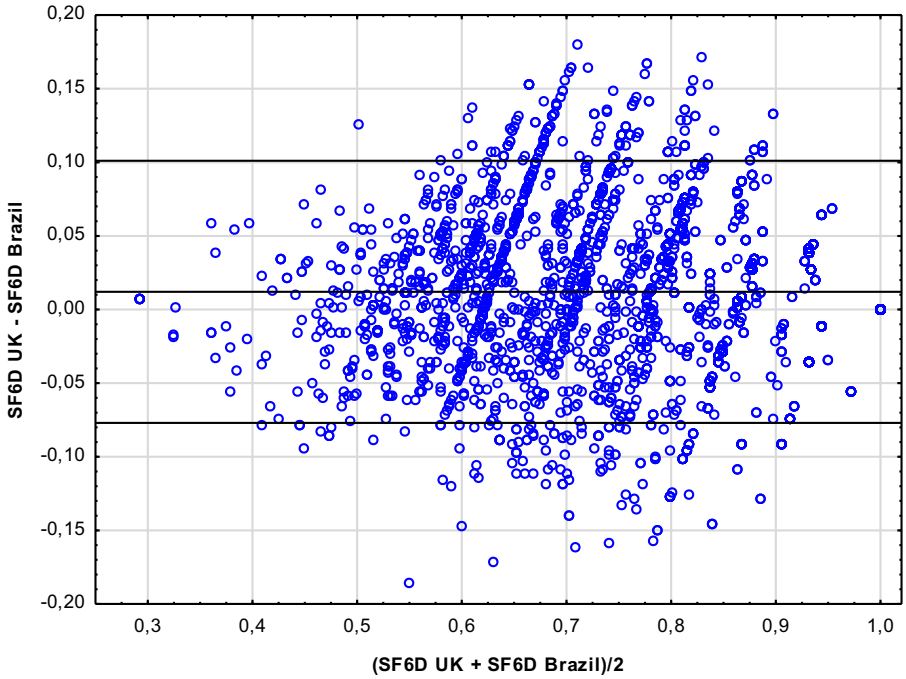


Fig. 2 – The Bland-Altman method: plot of the difference between SF-6D Brazil and SF-6D UK against the mean of the two SF-6D utility scores. The bias of 0.012 is represented by the gap between the X axis corresponding to zero difference and the X axis at 0.012 units. Limits of agreement (the range of values where 95% of differences between the two SF-6D utility scores are expected to be: −0.077 [CI: −0.079 to −0.075] to 0.101 [CI: 0.099 to 0.103]).

Table 3 – SF-6D values for a sample of 30 health states using Brazilian and UK preference weights.

Health state	SF-6D Brazil utilities	SF-6D UK utilities
111111	1.000	1.000
111112	0.950	0.967
211111	0.920	0.949
113112	0.917	0.927
111121	0.913	0.953
111211	0.911	0.94
111511	0.884	0.926
114115	0.848	0.888
221111	0.847	0.901
132121	0.840	0.867
122411	0.838	0.854
111512	0.834	0.893
211311	0.831	0.889
111221	0.824	0.893
611111	0.817	0.904
222112	0.797	0.830
211123	0.783	0.869
111243	0.773	0.860
133144	0.756	0.832
112624	0.717	0.795
611123	0.680	0.824
442344	0.606	0.723
333323	0.588	0.721
344344	0.540	0.701
424452	0.511	0.681
444554	0.470	0.667
445655	0.392	0.627
634555	0.345	0.603
645653	0.325	0.601
645655	0.289	0.582

SF-6D indicates 6-dimensional health state short form.

Table 4 – Brazilian and UK consistent models.

Domain/ level	Brazilian consistent model	UK consistent model
C	1.000	1.000
PF2	−0.080	−0.053
PF3	−0.080	−0.011
PF4	−0.080	−0.040
PF5	−0.080	−0.054
PF6	−0.183	−0.111
RL2	−0.073	−0.053
RL3	−0.073	−0.055
RL4	−0.087	−0.050
SF2	−0.033	−0.055
SF3	−0.033	−0.067
SF4	−0.066	−0.070
SF5	−0.078	−0.087
PAIN2	−0.089	−0.047
PAIN3	−0.089	−0.025
PAIN4	−0.089	−0.056
PAIN5	−0.116	−0.091
PAIN6	−0.146	−0.167
MH2	−0.087	−0.049
MH3	−0.087	−0.042
MH4	−0.088	−0.109
MH5	−0.131	−0.128
VIT2	−0.050	−0.086
VIT3	−0.050	−0.061
VIT4	−0.050	−0.054
VIT5	−0.086	−0.091
Most	–	−0.070
Least	–	–

C indicates constant; MH, mental health; PF, physical functioning; RL, role limitation; SF, social functioning; VIT, vitality.

used as a health index; (4) explore the variability in health preferences according to social and cultural characteristics of the population; and (5) test whether the respective magnitudes of different SF-6D scoring formulas would result in a greater likelihood of accepting or not the incremental cost-utility ratio.

It is also important to consider that the use of national preference weights from HRQOL studies seems to be more relevant and appropriate than the use of UK preference weights, because they reflect cultural and country-specific preferences for health states.[12,14,25] In this sense, the description of the results in terms of overall means does not allow the discrimination of the importance that each culture considers for multiple health dimensions. So, even though there were no differences between the values of health states using the Brazilian or UK algorithm, based exclusively on means according to sociodemographic variables, the SF-6D weights (model coefficients) can differ, as presented in Table 3. Nevertheless, data presented in the present work, showing high correlations and agreement between the current Brazilian and UK SF-6D weight system, raise doubts about the importance of these efforts.

Study Limitations

This study had limitations. The sample was not representative of rural areas and the findings may not be applicable to subjects younger than 15 years, who were excluded from the study. Furthermore, the association between cross-cultural variables and preference measurement, and how this could have an impact on the micro-utility effects, was not explored. We also did not consider that within-country variations may be as large as or larger than between-country variations in preference structures. Furthermore, we also did not consider that we would potentially expect greater differences to occur between Brazilian and UK value sets for people with more severe health rather than at the upper end of health, where utilities will be close to 1. In addition, health states defined by the Brazilian SF-6D system were valued only by a sample of the Southern Brazilian population. It is also important to consider that this may not reflect the preferences of the Brazilian population as a whole. This would be of great significance in Brazil, a country of continental dimensions with socioeconomic differences in their various regions, thus highlighting the importance of an assessment that is representative of the different regions of the country.[39]

Methodological aspects were also not considered, such as order and framing effects of descriptions in valuation tasks in Brazil and in the United Kingdom, the influence of different regression models, variability in health preferences by respondent characteristics, and time of survey completion.

Although in theory the development of country-specific system weights for SF-6D (or other preference-based measures of health) makes sense, in spite of being time-consuming and resource-intensive, this study raises certain questions to be further evaluated in futures studies: (1) Is one sample alone of a defined population (that may not represent the whole country population) an adequate sample to express and reflect the population preference? (2) Can the differences observed when using different country-specific system weights be considered better than the MID to justify and recognize the value of a healthcare

intervention? (3) Are the differences observed using different country-specific system weights greater than the differences reported by test-retest reliability studies of preference-based outcome measures? and (4) Do the development and use of country-specific system weights (as an outcome measure itself, or as part of economic evaluation) influence and generate distinct individual and collective healthcare decisions?

Conclusions

The present study identified quantitatively small differences between UK- and Brazilian-derived SF-6D scores. The relevance of these differences may, however, be questioned on the basis of tests of agreement. Nevertheless, comparisons between the results of different countries' national weights are important to enable users to fully understand the impact of using different sets of preference weights so that QALY can be used as a scientifically sound metric to guide healthcare decisions.

Acknowledgments

Source of financial support: This study was partially funded by the Brazilian Health Ministry—Cooperation Term Federal University of São Paulo (no. 136/2011).

REFERENCES

1. Geisler BP. Health state survey-derived utilities in cost-utility analysis—a call to action. Value Health 2010;13(2):159.
2. Wisloff T, Hagen G, Hamidi V, et al. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics 2014;32(4):367–75.
3. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013. London: National Institute for Health and Care Excellence; 2014.
4. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA 2016;316(10):1093–103.
5. Arnold D, Girling A, Stevens A, Lilford R. Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. BMJ 2009;339:b2688.
6. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21(2):271–92.
7. Brazier J, Connell J, Papaioannou D, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess 2014;18(34):1–188.
8. Konerding U. Preference-based index measurement of health-related quality of life: when does it reflect only arbitrary settings of the researcher? Health Econ 2011;20(4):471–83.
9. Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002;11(4):341–53.
10. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: a call to reconsider current guidelines. Soc Sci Med 2016;165:66–74.
11. Prieto L, Sacristan JA. What is the value of social values? The uselessness of assessing health-related quality of life through preference measures. BMC Med Res Methodol 2004;4:10.
12. Cruz LN, Camey SA, Hoffmann JF, et al. Estimating the SF-6D value set for a population-based sample of Brazilians. Value Health 2011;14(5 suppl 1):S108–14.
13. Ferreira LN, Ferreira PL, Pereira LN, Brazier J, Rowen D. A Portuguese value set for the SF-6D. Value Health 2010;13(5):624–30.
14. Brazier JE, Fukuhara S, Roberts J, et al. Estimating a preference-based index from the Japanese SF-36. J Clin Epidemiol 2009;62(12):1323–31.
15. Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. Med Decis Making 2013;33(6):793–803.
16. Wee HL, Cheung YB, Fong KY, et al. Are English- and Chinese-language versions of the SF-6D equivalent? A comparison from a population-based study. Clin Ther 2004;26(7):1137–48.
17. Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. Value Health 2008;11(2):295–303.
18. Campolina AG, Bortoluzzo AB, Ferraz MB, Ciconelli RM. Validation of the Brazilian version of the generic six-dimensional short form quality of life questionnaire (SF-6D Brazil). Cien Saude Colet 2011;16(7):3103–10.
19. Fleiss JL. The Design and Analysis of Clinical Experiments. Chichester, UK: Wiley Classics Library; 1999.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–10.
21. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141–51.
22. Huang IC, Willke RJ, Atkinson MJ, et al. US and UK versions of the EQ-5D preference weights: does choice of preference weights make a difference? Qual Life Res 2007;16(6):1065–72.
23. Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? Med Care 2005;43(3):221–8.
24. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21(1):7–16.
25. Ferreira LN, Ferreira PL, Rowen D, Brazier JE. Do Portuguese and UK health state values differ across valuation methods? Qual Life Res 2011;20(4):609–19.
26. Wong CK, Lam ET, Lam CL. Comparison of direct-measured and derived short form six dimensions (SF-6D) health preference values among chronic hepatitis B patients. Qual Life Res 2013;22(10):2973–81.
27. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. J Shoulder Elbow Surg 2010;19(3):342–8.
28. Tran BX, Nguyen LH, Ohinmaa A, et al. Longitudinal and cross sectional assessments of health utility in adults with HIV/AIDS: a systematic review and meta-analysis. BMC Health Serv Res 2015;15:7.
29. Khanna D, Furst DE, Wong WK, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. Qual Life Res 2007;16(6):1083–92.
30. Palta M, Chen HY, Kaplan RM, et al. Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. Med Decis Making 2011;31(2):260–9.
31. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420–8.
33. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim JS. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. Med Decis Making 2010;30(1):5–15.
34. Whitehurst DG, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? Value Health 2011;14(4):531–8.
35. Richardson J, Iezzi A, Khan MA. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and "micro-utility" effects. Qual Life Res 2015;24(8):2045–53.
36. Richardson J, Khan MA, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. Med Decis Making 2015;35(3):276–91.
37. Ferreira LN, Ferreira PL, Pereira LN, Rowen D, Brazier JE. Exploring the consistency of the SF-6D. Value Health 2013;16(6):1023–31.
38. Davis JC, Liu-Ambrose T, Khan KM, Robertson MC, Marra CA. SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. Osteoporos Int 2012;23(7):1849–57.
39. Ciconelli RM, Ferraz MB, Kowalski S, da Rocha Castelar Pinheiro G, Sato EI. Brazilian urban population norms derived from the health-related quality of life SF-6D. Qual Life Res 2015;24(10):2559–64.
40. Instituto Brasileiro de Geografia e Estatística. Censo 2010. http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default_resultados_universo.shtm. Accessed May 3, 2015.