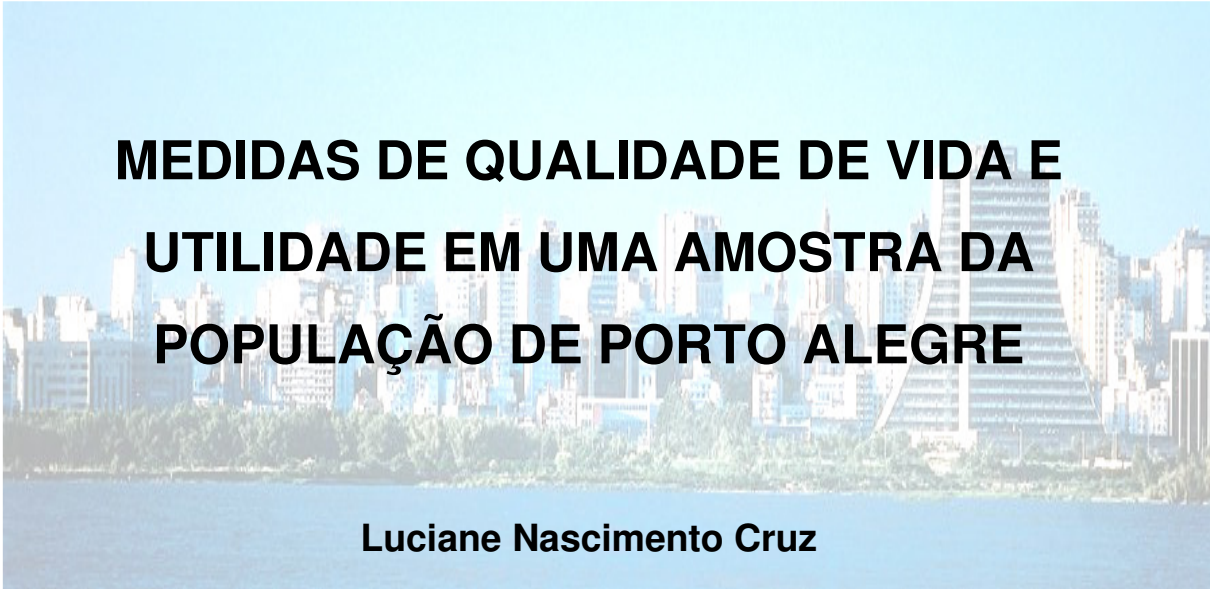


**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**

TESE DE DOUTORADO



**MEDIDAS DE QUALIDADE DE VIDA E
UTILIDADE EM UMA AMOSTRA DA
POPULAÇÃO DE PORTO ALEGRE**

Luciane Nascimento Cruz

Porto Alegre, Brasil

2010

C957m Cruz, Luciane Nascimento

Medidas de qualidade de vida e utilidade em uma amostra da população de Porto Alegre / Luciane Nascimento Cruz ; orient. Carisi Anne Polanczyk ; co-orient. Marcelo Pio de Almeida Fleck. – 2010.
270 f. : il. color.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul.
Faculdade de Medicina. Programa de Pós-Graduação em Epidemiologia.
Porto Alegre, BR-RS, 2010.

1. Qualidade de vida 2. Porto Alegre (RS) 3. Assistência à saúde 4.
Avaliação de custo-efetividade 5. Epidemiologia I. Polanczyk, Carisi Anne
II. Fleck, Marcelo Pio de Almeida III. Título.

NLM: WA 30

Catálogo Biblioteca FAMED/HCPA

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



TESE DE DOUTORADO

**MEDIDAS DE QUALIDADE DE VIDA E
UTILIDADE EM UMA AMOSTRA DA
POPULAÇÃO DE PORTO ALEGRE**

Luciane Nascimento Cruz

A apresentação desta tese é exigência do Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Doutor.

Orientador: Profa. Dra. Carisi Anne Polanczyk

Co-orientador: Prof. Dr. Marcelo Pio de Almeida Fleck

Porto Alegre, Brasil

2010

BANCA EXAMINADORA – Defesa final

Prof. Dr. Federico Augustovski

Escola de Saúde Pública
Faculdade de Medicina
Universidade de Buenos Aires

Prof. Dra. Cristiana Maria Toscano Soares

Pós-doutoranda
Programa de Pós-Graduação em Epidemiologia
Universidade Federal do Rio Grande do Sul
Professora colaboradora
Universidade Federal de Goiás

Prof. Dr. Álvaro Vigo

Programa de Pós-Graduação em Epidemiologia
Universidade Federal do Rio Grande do Sul

BANCA EXAMINADORA – Defesa preliminar

Prof. Dr. Bruce Bartholow Duncan

Programa de Pós-Graduação em Epidemiologia
Universidade Federal do Rio Grande do Sul

Prof. Dra. Daniela Riva Knauth

Programa de Pós-Graduação em Epidemiologia
Universidade Federal do Rio Grande do Sul

Prof. Dr. Álvaro Vigo

Programa de Pós-Graduação em Epidemiologia
Universidade Federal do Rio Grande do Sul

Esta tese é dedicada:

*Aos meus pais, **Valmi** e **Sérgio**, que me ensinaram que estudar era a melhor coisa pra “ficar forte quando crescer”...*

*Ao **Fernando**, que acrescentou qualidade à minha vida.*

*E àquela “galera de casa” que faz a “alegria do dia-a-dia”, meu irmão **Duda**, a **Lê**, a fofíssima **Lara** e a tia **Malu**.*

Agradecimento especial

- À minha orientadora, **Prof. Dra. Carisi Anne Polanczyk**, a quem devo minha identidade de pesquisadora, as oportunidades de participar da geração de conhecimento e a chance de fazer disso uma atividade profissional. Com seu carisma, competência e “inteligência social” possibilita crescimento a todos que compartilham do seu trabalho.
- Ao meu co-orientador, **Prof. Dr. Marcelo Pio de Almeida Fleck**, por ter me inserido neste universo da qualidade de vida e possibilitado meu aprimoramento neste campo através de seu sólido conhecimento, rigor metodológico e capacidade de reconhecer os avanços e mostrar o que ainda pode ser feito.
- À **Prof. Dra. Suzi Alves Camey**, que na verdade exerceu um papel de “co-co-orientadora”, sem a qual o caminho para atingir os resultados deste trabalho teria sido duríssimo, se não impossível.

Agradecimentos

- À minha “super” coordenadora de campo, agora bióloga, **Michele Rosana Oliveira**, pela dedicação ao meu projeto como se fosse o seu próprio e por ter sido responsável pelo rigor na coleta de dados.
- À minha grande colaboradora **Juliana Hoffmann**, que participou deste trabalho de maneira “multiuso”, desde a formação do banco de dados, análises e apresentação dos resultados.
- Às colegas **Cristiane Melere** e **Mônica Oliveira** por terem facilitado minha “relação” com o *software* Teleform e pelas ajudas *express* sempre que precisei.
- Aos **Profs. Drs. Erno Harzheim** e **Maria Angélica Nunes**, pela disponibilidade em me oferecer consultoria metodológica no planejamento da amostra e coleta de dados.
- A **toda minha equipe de entrevistadoras**, pelo fôlego de andar pelas ruas de Porto Alegre, com sol e com chuva em busca dos dados.

- Aos **Profs. John Edward Brazier** e **Aki Tsuchiya**, as pesquisadoras **Catherine Stevens** e **Donna Rowen**, da Universidade de Sheffield, Inglaterra, pela colaboração no planejamento do projeto, treinamento para coleta de dados, análises dos resultados e revisão dos artigos.
- Aos **colegas do grupo de qualidade de vida do PPG Psiquiatria**, pelas constantes contribuições durante as reuniões.
- À colega e amiga **Ana Flávia Lima**, pela ajuda de diversas formas prestada neste final de doutorado.
- Ao **PPG Epidemiologia**, representado pelos **Profs. Drs. Sotero Mengue, Bruce Duncan** e **Maria Inês Schmidt** que me acolheram como uma “filha pródiga” lá no mestrado e desde então eu não deixei mais esta família.
- Às colegas “epidemiológicas” e parceiras de *happy hour* **Suzi Camey, Luciana Nunes, Helena Santos** e **Patrícia Manzolli** pela “força moral”.
- Às parceiras de sala **Mariana Furtado** e **Juliana Juk**, pelos auxílios logísticos neste último ano e pela agradável convivência.
- Aos **Profs. Bruce Bartholow Duncan, Daniela Riva Knauth** e **Álvaro Vigo** pelas preciosas contribuições na banca preliminar que possibilitaram o aprimoramento científico do meu trabalho.
- À **minha família**, tanto a “de sangue” quanto a “adquirida” pelo apoio incondicional.
- À **Dra. Magda Lia Teixeira Martins** por ajudar a manter a minha “mente quieta, a espinha ereta e o coração tranquilo”.
- À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)** pelo auxílio financeiro ao programa de doutorado.
- Ao **Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)** pelo financiamento do projeto.
- Ao **Fundo de Incentivo a Pesquisa e Eventos (FIPE)** do Hospital de Clínicas de Porto Alegre pelo suporte financeiro para tradução dos artigos.
- A **todos os moradores de Porto Alegre** que abriram a porta de suas casas para colaborar com o projeto
- E temendo esquecer alguém, muito obrigada **a todos que contribuíram com este trabalho**, porque ciência não se faz sozinho, e a colaboração é a chave para a evolução.

SUMÁRIO

ABREVIATURAS E SIGLAS

RESUMO

ABSTRACT

LISTA DE TABELAS

LISTA DE FIGURAS

APRESENTAÇÃO

1 INTRODUÇÃO	20
2 REVISÃO DA LITERATURA	24
2.1 QUALIDADE DE VIDA	24
2.1.1 Conceito de qualidade de vida	24
2.1.2 Instrumentos para medir qualidade de vida	29
2.1.2.1 Características de um bom instrumento	31
2.1.2.2 Tipos de instrumentos	33
2.1.3 Interpretação dos escores das medidas de qualidade de vida	37
2.1.4 Instrumentos genéricos de avaliação de qualidade de vida: SF-36 e WHOQOL-breve	40
2.1.4.1 <i>Medical Outcomes Study Short-Form 36 (SF-36)</i>	40
2.1.4.2 <i>World Health Organization Quality of Life Instrument-bref</i> (WHOQOL-breve)	46
2.1.5 Medidas de utilidade	50
2.1.5.1 Base teórica	51
2.1.5.2 Desenvolvimento das medidas	54
2.1.5.3 Avaliação dos métodos baseados em preferências	75
2.1.5.4 Efeitos da população estudada e do contexto de medida nos índices de utilidade	78
2.2 APLICAÇÃO PRÁTICA DAS MEDIDAS DE UTILIDADE	81
3 OBJETIVOS	86
3.1 OBJETIVOS PRINCIPAIS	86
3.2 OBJETIVOS SECUNDÁRIOS	86

4 REFERÊNCIAS BIBLIOGRÁFICAS	88
5 ARTIGOS	
ARTIGO 1: Estimativa de índices de utilidade a partir do SF-36 no Brasil: descrição da metodologia	95
ARTIGO 2: Inconsistencies and SF-6D health states valuation in Brazil	120
ARTIGO 3: Estimating the SF-6D value set for a population based sample of Brazilians	146
ARTIGO 4: Health-related quality of life in Brazil: normative data for the SF-36 in a southern general population sample	175
ARTIGO 5: Quality of life in Brazil: normative values for the WHOQOL- Bref in a southern general population sample	201
6 CONCLUSÕES E CONSIDERAÇÕES FINAIS	224
7 ANEXOS	226
ANEXO 1 - Projeto de Pesquisa	
ANEXO 2 - Termo de Consentimento Informado	
ANEXO 3 - Protocolo do Estudo	

Abreviaturas e Siglas

DALY	<i>Disability-Adjusted Life-Years</i>
DP	Desvio-Padrão
ES	<i>Effect Size</i>
EQ-5D	<i>Euro Quality of Life Instrument-5D</i>
EAV	Escala Análogo Visual
HUI	<i>Health Utility Index</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IQOLA	<i>International Quality of Life Assessment Project</i>
MAU	<i>Multiattribute Utility Theory</i>
NICE	<i>National Institute of Clinical Excellence</i>
OMS	Organização Mundial da Saúde
PIB	Produto Interno Bruto
PRO	<i>Patient Reported Outcomes</i>
QALY	<i>Quality Adjusted Life Years</i>
QV	Qualidade de Vida
QWB	<i>Quality of Well-Being</i>
SAQ	<i>Seattle Angina Questionnaire</i>
SF-36	<i>Medical Outcomes Study Short-Form 36</i>
SF-6D	<i>Short-Form 6D</i>
SG	<i>Standard Gamble</i>
SRM	<i>Standardized Response Mean</i>
TRI	Teoria de Resposta ao Item
TTO	<i>Time trade-off</i>
WHOQOL	<i>World Health Organization Quality of Life Instrument</i>

RESUMO

Contexto: As medidas para avaliação de qualidade de vida (QV) vêm sendo cada vez mais utilizadas como um desfecho em ensaios clínicos e como medida de efetividade e de qualidade dos cuidados em saúde. Recentemente, são recomendadas por diretrizes internacionais e por órgãos internacionais de avaliação de tecnologias em saúde como um desfecho a ser utilizado em estudos de custo-efetividade. No cenário brasileiro, instrumentos traduzidos e validados na população estão sendo aplicados em vários grupos de pacientes para avaliar o impacto de diversas doenças e intervenções na vida dos indivíduos afetados. Contudo, a falta de escores normativos de qualidade de vida, baseados na população geral brasileira, que representem um referencial contra o qual dados de outros subgrupos possam ser comparados, dificulta a interpretação dos resultados dos estudos.

Além disso, há no Brasil uma carência de instrumentos de qualidade de vida adequados para utilização em análises econômicas. Portanto, a adaptação cultural destes instrumentos no contexto nacional torna-se tarefa imprescindível no desenvolvimento e implantação de metodologias de estudos em economia em saúde no Brasil.

Objetivos: Disponibilizar instrumentos de qualidade de vida para uso como medida de desfecho em análises econômicas e estudos clínicos e epidemiológicos, através de duas estratégias principais: 1) Realizar a adaptação cultural do instrumento britânico SF-6D, adequado para uso em análises econômicas; 2) Descrever dados normativos populacionais para os escores dos instrumentos genéricos SF-36 e WHOQOL-breve que poderão ser utilizados como referência para comparações entre grupos.

Métodos: Uma amostra da população geral adulta de Porto Alegre com idade entre 20 e 64 anos de idade foi randomicamente selecionada através dos setores censitários do município divididos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Para a coleta de dados, o seguinte protocolo foi aplicado aos indivíduos participantes: a) Preenchimento dos itens do SF-6D; b) ordenamento de 8 estados de saúde hipotéticos gerados pelo sistema descritivo do SF-6D; c) aplicação da técnica *standard gamble* para obtenção dos valores das preferências dos participantes pelos estados de saúde; d) preenchimento dos questionários SF-36 e

WHOQOL-breve; e) questionário sociodemográfico. As entrevistas foram realizadas no domicílio dos participantes por uma equipe treinada de entrevistadores.

Modelos de regressão foram estimados utilizando os dados obtidos através da aplicação do *standard gamble*. O modelo com melhor desempenho foi utilizado para construção do algoritmo SF-6D para o Brasil, possibilitando estimar índices de utilidade a partir do instrumento genérico SF-36. Estatística descritiva dos escores do SF-36 e WHOQOL-breve foi realizada de acordo com características socio-demográficas.

Resultados: Um total de 528 indivíduos participou do protocolo de medida de preferências através da técnica *standard gamble*. Na avaliação das respostas obtidas, 28% dos valores foram considerados inconsistentes. A avaliação das características socio-demográficas demonstrou que o grupo com respostas inconsistentes era formado por sujeitos mais velhos ($p = 0,018$), de menor nível educacional ($\chi^2 = 19,90$ $p < 0,01$) e de menor classe econômica ($\chi^2 = 16,09$ $p = 0,007$). Dados dos 382 participantes com respostas consistentes ao *standard gamble* foram utilizados para estimar modelos preditores utilizados para construção de um algoritmo para cálculo dos escores do SF-6D. O modelo com melhor ajuste aos dados foi um modelo de efeitos randômicos contendo apenas as variáveis de efeitos principais, diferente do modelo preconizado para cálculo do SF-6D no Reino Unido, salientando a importância de utilizar um algoritmo específico para cada país.

Em relação aos resultados obtidos pela aplicação dos instrumentos psicométricos a 755 indivíduos, tanto a qualidade de vida relacionada à saúde, medida pelo SF-36, como a qualidade de vida mais global, medida pelo WHOQOL-breve, apresentou-se reduzida nas mulheres, em sujeitos de classe econômica mais baixa, de menor nível educacional e com auto-relato de condição médica crônica. Escores normativos de acordo com sexo e faixa etária são apresentados para os quatro domínios do WHOQOL-breve e os oito domínios do SF-36.

Conclusão: Os resultados da aplicação do *standard gamble* fornecem a primeira base de dados populacional para preferências por estados de saúde no Brasil, possibilitando o cálculo de QALYs para estudos de custo-utilidade utilizando dados regionais. Escores normativos para o SF-36 e o WHOQOL-breve podem ser utilizados para comparações de resultados obtidos em grupos de pacientes. A habilidade para discriminar diferentes subgrupos populacionais torna estas medidas ferramentas úteis na identificação de grupos vulneráveis em estudos epidemiológicos.

ABSTRACT

Background: Quality of Life (QOL) measures has been used as an outcome in clinical trials and as a measure of effectiveness and quality of health care. Recently, QOL has been recommended by international guidelines and international institutions of Health Technology Assessment such as the National Institute of Clinical Excellence (NICE), in the United Kingdom, as an outcome to be used in studies of cost-effectiveness. In the Brazilian scenario, instruments translated and validated for the Brazilian population are being used in different patient groups to assess the impact of different diseases and interventions on the lives of affected individuals. However, the lack of QOL normative scores based on the general Brazilian population, representing a reference against data from other subgroups can be compared, makes the interpretation of studies results more difficult.

Moreover, in Brazil preference-based measures of quality of life are lacking, avoiding generating utility values for using in economic analysis. Therefore, the cultural adaptation of this kind of instruments in the national context becomes an essential task in the development and implementation of methodologies for health economics in Brazil.

Objectives: To provide quality of life measures to use as an outcome in economic analysis, clinical and epidemiological studies through two main strategies: 1) To measure health state values based on the British preference-based instrument called SF-6D, using the values for the construction of an algorithm to estimate utility values for application in the national context; 2) To provide data for the normative population scores of generic instruments SF-36 and WHOQOL-brief, describing values that can be used as a reference for comparisons between groups.

Methods: A sample of the adult general population of Porto Alegre aged from 20 to 64 was randomly selected through census sectors divided by the Brazilian Institute of Geography and Statistics (IBGE). For the data collection, the following standard was applied: a) Completion of the items of the SF-6D; b) a ranking exercise using 8 hypothetical health states generated by the SF-6D descriptive system; c) application of the standard gamble technique to obtain the values of preferences for health states d) completion of the SF-36 and WHOQOL-brief e) socio-demographic

questionnaire. Regression models were estimated using the data obtained through the preferences for health states using standard gamble. The best fit model was used for construction of the SF-6D algorithm for Brazil, making it possible to estimate utility values from the SF-36. Descriptive statistics of the scores of the SF-36 and WHOQOL-brief was carried out according to socio-demographic characteristics.

Results: Data from 528 individuals were available to analyze preferences for health states using standard gamble technique. In evaluating the responses obtained, 28% of values were considered inconsistent. The evaluation of socio-demographic characteristics showed that the group with inconsistent responses were older ($p = 0.018$), less educated ($\chi^2 = 19.90$ $p < 0.01$) and from a lower income group ($\chi^2 = 16.09$ $p = 0.007$). Data from 382 participants with consistent responses were used to estimate predictive models used to construct an algorithm for calculating the scores of the SF-6D. The model with best fit to the data was a random effects model containing only variables of main effects, unlike the model recommended for calculating the SF-6D in the UK, highlighting the importance of using a specific algorithm for each country.

Related to results obtained through application of SF-36 and WHOQOL-bref, both the health-related quality of life, as measured by SF-36, and the general quality of life, as measured by the WHOQOL-bref, presented reduced in women, in subjects from a lower economic class, less educated and self-reporting chronic medical conditions. Normative scores by gender and age are presented for the 4 domains of WHOQOL-BREF and the 8 domains of the SF-36 in the form of tables and charts percentiles.

Conclusion: The results of applying the standard gamble provide the first population-based data set for preferences for health states in Brazil, enabling the calculation of QALYs for cost-utility analysis using regional data. Normative scores described to both the SF-36 and for the WHOQOL-bref may be useful for comparisons of results obtained in groups of patients. The ability to discriminate between different population subgroups makes such measures useful tools in identifying vulnerable groups in epidemiological studies.

Lista de Tabelas

Tabela 1	Tipos de instrumentos para avaliação de qualidade de vida.....	35
Tabela 2	Taxonomia do SF-36.....	42
Tabela 3	Sumário do conteúdo do SF-36.....	44
Tabela 4	Valores obtidos e reprodutibilidade para cada domínio do questionário SF-36 e o intervalo de tempo entre as duas aplicações na validação para população brasileira com artrite reumatóide.....	46
Tabela 5	Domínios e facetas do WHOQOL-breve.....	48
Tabela 6	Análise de variância do domínio em relação a normal X paciente, coeficiente de fidedignidade de Cronbach dos domínios e questões e coeficiente de correlação dos escores dos diferentes domínios entre teste e reteste na aplicação da versão em português do WHOQOL-breve.....	49
Tabela 7	Métodos para medir preferências.....	53
Tabela 8	Confiabilidade dos métodos de medida direta de utilidade.....	76

Lista de Figuras

Figura 1	Exemplo de um estado de saúde “multiatributo”	56
Figura 2	Teoria do processamento de informação.....	59
Figura 3	Sistema descritivo do instrumento EuroQol-5D	62
Figura 4	Exemplo de estado de saúde gerado pelo EuroQol-5D.....	63
Figura 5	Representação gráfica do <i>standard gamble</i>	68
Figura 6	Representação gráfica do <i>standard gamble</i> para estados piores que a morte	69
Figura 7	Representação gráfica do <i>standard gamble</i> para estados temporários	70
Figura 8	Representação gráfica do <i>Time trade-off</i>	71

APRESENTAÇÃO

Este trabalho consiste na tese de doutorado intitulada “**Medidas de qualidade de vida e utilidade em uma amostra da população de Porto Alegre**”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 11 de junho de 2010.

O estudo que originou a presente tese é parte do “Projeto Colaborativo Interinstitucional de Estudos de Custo-efetividade em Saúde”, contemplado pelo Edital do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), inserido na linha de pesquisa de Avaliação econômica e análise de custos em saúde (Edital MCT-CNPq / MS-SCTIE-DECIT – Nº 36/2005). O objetivo geral do projeto foi consolidar uma linha de pesquisa, sua correspondente plataforma e infra-estrutura necessárias para a realização de estudos, pesquisas e formação de recursos humanos em análises econômicas e avaliação de tecnologias em saúde aplicadas a população brasileira, com ênfase em ações e cenários no âmbito do Sistema Único de Saúde. O projeto contou com a participação de pesquisadores oriundos de vários Programas de Pós-Graduação (PPG) da Universidade Federal do Rio Grande do Sul (Epidemiologia, Cardiologia, Psiquiatria e Engenharia de Produção) e do Hospital de Clínicas de Porto Alegre e também de colaboradores das instituições internacionais *School of Health and Related Research* (SChARR) da Universidade de Sheffield, na Inglaterra e Organização Pan-Americana da Saúde, por parte da Unidade de Prevenção e Controle de Doenças (*Center of Disease Control (CDC) Diabetes Study Group*).

O grupo que esteve mais diretamente envolvido com o estudo apresentado nesta tese foi formado por professores e pesquisadores dos PPGs de Epidemiologia e Psiquiatria da Universidade Federal do Rio Grande do Sul e os colaboradores da Universidade de Sheffield.

O trabalho é apresentado em três partes, na ordem que segue:

- 1) Introdução, Revisão da Literatura e Objetivos
- 2) Artigo(s):
 - Artigo 1: Estimativa de índices de utilidade a partir do SF-36 no Brasil:

descrição da metodologia

- Artigo 2: Inconsistencies and SF-6D health states valuation in Brazil
- Artigo 3: Estimating the SF-6D value set for a population based sample of Brazilians
- Artigo 4: Health-related quality of life in Brazil: normative data for the SF-36 in a Southern general population sample
- Artigo 5: Quality of Life in Brazil: Normative Values for the WHOQOL-Bref in a Southern general population sample

3) Conclusões e Considerações Finais.

Documentos de apoio, incluindo o Projeto de Pesquisa, estão apresentados nos anexos.

1 INTRODUÇÃO

1 INTRODUÇÃO

A importância da avaliação de qualidade de vida (QV) aumentou de modo expressivo durante os últimos 50 anos. Esta maior relevância deve-se principalmente por ser um enfoque que valoriza a perspectiva do paciente e permite avaliar o real impacto de uma enfermidade e objetivos de seu tratamento sob um aspecto multidimensional, ou seja, além de simplesmente levar em conta a redução da morbidade e mortalidade (FLECK, 2000). Tal multidimensionalidade possibilita a abordagem da saúde dos indivíduos em diferentes domínios como, por exemplo, aspectos físicos, funcionamento no dia-a-dia, desempenho social e aspectos emocionais.

A medida de QV é imperativa, uma vez que insere o indivíduo nas avaliações relacionadas à sua própria saúde e tratamentos. Neste contexto, esta medida vem sendo cada vez mais utilizada como um desfecho em ensaios clínicos e como medida de efetividade e de qualidade dos cuidados em saúde. Fatores que têm contribuído para o uso crescente incluem o acúmulo de evidências de que ela seja uma medida válida e confiável, a publicação de grandes estudos clínicos demonstrando que estas medidas são responsivas a mudanças clínicas e o desenvolvimento de instrumentos mais curtos de mais fácil administração e compreensão (WILSON & CLEARY, 1995).

Independente da maneira pela qual a QV é avaliada, na pesquisa em saúde o propósito deste tipo de avaliação é, invariavelmente, o de medir o impacto funcional e subjetivo das doenças crônicas e seu tratamento na vida dos indivíduos afetados. E este é um dos objetivos mais importantes, uma vez que uma mesma doença pode ter efeitos diferentes em pessoas diferentes. Medidas fisiológicas fornecem informação para os clínicos, mas podem ser de interesse limitado para os pacientes (GUYATT, FEENY, PATRICK, 1993).

Embora as medidas objetivas para controle das doenças sejam essenciais, como dados laboratoriais, por exemplo, a perspectiva do paciente garante uma ampla compreensão dos efeitos da doença e de seu tratamento na sua vida (WILSON & CLEARY, 1995).

A importância da medida de QV tornou-se ainda mais evidente com a incorporação mais recente deste tipo de desfecho nas avaliações de tecnologias em saúde e nos processos de tomada de decisão em políticas públicas de saúde. Cada vez mais os gestores, provedores, pacientes e o público exigem que cada gasto adicional em saúde seja justificado de acordo com desfechos esperados. Dentro deste cenário, a tomada de decisão em saúde nunca foi tão importante, para reduzir ineficiência, eliminar procedimentos médicos não efetivos, aumentar a competitividade, melhorar a qualidade, modificar as fórmulas de reembolso e racionalizar os serviços (PATRICK & ERICKSON, 1996). Considerando que a sociedade não está disposta a aceitar um aumento dos gastos nos cuidados em saúde, as decisões deverão ser feitas utilizando estimativas dos custos e dos benefícios de cada alternativa. Nas últimas décadas, a literatura vem advogando o uso da QV como uma medida dos benefícios trazidos pelos gastos em saúde e para avaliar a estrutura e o processo da assistência à saúde de uma população. Alguns autores enfocam a QV relacionada à saúde como a mais relevante e abrangente medida de desfecho para comparar custos (PATRICK & ERICKSON, 1996). Neste contexto, a medida de QV vem sendo recomendada como a medida de efetividade a ser utilizada em estudos de custo-efetividade por diretrizes internacionais (WEINSTEIN et al., 1996) e por órgãos internacionais de avaliação de tecnologias em saúde como o *National Institute of Clinical Excellence* (NICE), do Reino Unido (2010b).

O uso da QV em ensaios clínicos também tem sido fortemente recomendado, principalmente nas seguintes situações: a) como desfecho principal em ensaios avaliando cuidados paliativos; b) quando um novo tratamento é equivalente ao tratamento padrão em termos de eficácia, mas pode trazer benefícios na QV dos pacientes; c) quando uma nova intervenção aumenta as taxas de cura ou sobrevida, mas este efeito pode ser anulado por uma deterioração importante da QV (FAYERS & MACHIN, 2007).

Recente revisão do uso de instrumentos que avaliam desfechos auto-relatados (*Patient Reported Outcomes* – PRO), um conceito atual e mais amplo que inclui QV, demonstrou que o uso destas medidas em ensaios clínicos aumentou de 4% para 12% na década atual (SCOGGINS & PATRICK, 2009).

Alguns autores acreditam que a combinação de medidas de QV com indicadores clínicos e outros desfechos objetivos produzem um cenário mais abrangente na avaliação de uma tecnologia em saúde. Existem evidências de que os desfechos subjetivos e objetivos nem sempre são concordantes, ou seja, a experiência subjetiva de uma doença correlaciona-se muito pouco com o estado clínico dos pacientes. Portanto, o uso agregado de medidas subjetivas e objetivas garante uma avaliação mais fidedigna do resultado alcançado com as diferentes intervenções em saúde (MOSTSELLER & FALOTICO-TAYLOR, 1989).

No cenário brasileiro, a pesquisa em QV vem crescendo, e instrumentos genéricos de uso disseminado no mundo todo como o *Medical Outcomes Study Short-Form 36* (SF-36) (WARE JR. & SHERBOURNE, 1992) e *World Health Organization Quality of Life Instrument-breve* (WHOQOL-breve, 1998a) já estão traduzidos e validados para a população brasileira. Estas medidas vêm sendo aplicadas em várias populações de pacientes para avaliar o impacto de diversas doenças e intervenções na vida dos indivíduos afetados. Contudo, a falta de escores normativos de QV, baseados na população geral brasileira, que representem um “padrão-ouro” contra o qual dados de outros subgrupos possam ser comparados, dificulta a interpretação dos resultados dos estudos.

Além disso, há no Brasil uma carência dos chamados instrumentos de QV baseados em preferências, que geram índices de utilidade. Considerando as recomendações internacionalmente estabelecidas para a realização de estudos de análise econômica em saúde que ressaltam a necessidade de utilização de parâmetros de utilidade para a análise de custo-efetividade (WEINSTEIN et al., 1996), a falta destas medidas determina um grande hiato para a pesquisa em saúde no país. Portanto, a adaptação cultural de instrumentos que gerem índices de utilidade para a população brasileira torna-se tarefa imprescindível no desenvolvimento e implantação de metodologias de estudos em economia em saúde no Brasil.

2 REVISÃO DA LITERATURA

2 REVISÃO DA LITERATURA

2.1 QUALIDADE DE VIDA

2.1.1 Conceito de qualidade de vida

As crescentes inovações tecnológicas da medicina que vem ocorrendo nas últimas décadas trazem como consequência, entre outras, um aumento da expectativa de vida da população. Este aumento dos anos de vida ocorre não só para os indivíduos saudáveis, mas também para aqueles portadores de enfermidades crônicas, uma vez que os tratamentos disponíveis permitem maior controle de sintomas e um retardo do curso natural das doenças. Assim, passa a ser de grande importância disponibilizar maneiras de avaliar como as pessoas estão vivendo estes anos “ganhos”. A introdução do conceito de QV como uma medida de desfecho em saúde surgiu nesse contexto, a partir da década de 70, podendo-se identificar algumas vertentes que contribuíram para o desenvolvimento da base conceitual (FLECK, 2008):

- Estudos epidemiológicos sobre felicidade e bem-estar: Surgiram nos países desenvolvidos, a partir de 1960. O objetivo era pesquisar como as pessoas se sentiam consigo mesmas, seus medos e ansiedades, os problemas que enfrentam e as formas de lidar com eles, seus pontos fortes e recursos (GURIN, VEROFF, FELD, 1960). Em 1976, Campbell, Converse e Rodgers, realizaram um estudo para monitorar a QV dos norte-americanos, avaliando as condições objetivas e subjetivas da população. Eles constataram, entre outros achados, que esta relação parecia não ter um padrão linear.

- Busca de indicadores sociais: No início do século XX, houve o surgimento dos indicadores de riqueza e desenvolvimento, como o Produto Interno Bruto (PIB), renda *per capita*, Índice de Mortalidade Infantil. Mais recentemente, foi criado o Índice de Desenvolvimento Humano (IDH) que agregou aspectos de natureza social e cultural (saúde e educação) às tradicionais medidas econômicas (FLECK, 2008).

- Necessidade de desfechos alternativos em saúde, além dos tradicionais morbidade e mortalidade: A avaliação baseada em exames laboratoriais e clínicos continuará sempre sendo essencial, mas pode se tornar insuficiente quando o objetivo é avaliar o impacto das enfermidades nas diferentes áreas da vida dos pacientes (FLECK, 2008).

- Interesse crescente em estudar aspectos humanísticos que extrapolam o foco exclusivo na doença: pesquisa das características adaptativas positivas das pessoas como resiliência, esperança, sabedoria, espiritualidade; preocupação com o grau de satisfação dos usuários em relação aos serviços prestados por determinada instituição, empresa ou serviço de saúde; reconhecimento da relação médico-paciente como responsável pelo sucesso das intervenções em saúde (FLECK, 2008).

Qualidade de vida é um termo ainda pouco definido. Na ausência de uma definição universalmente aceita, alguns autores argumentam que a maioria das pessoas, pelo menos no mundo ocidental, é familiar com a expressão “qualidade de vida” e teriam uma compreensão intuitiva do que o termo significa. Porém, fica claro que “qualidade de vida” pode significar coisas diferentes para pessoas diferentes, e recebe significados diversos dependendo da área de aplicação. Para uma área como planejamento urbanístico, por exemplo, pode representar o acesso a espaços arborizados, parques, etc. (FAYERS & MACHIN, 2007). É um termo que vem sendo utilizado em sociologia, psicologia, geografia, economia, história, medicina, farmácia, educação, criminologia, arquitetura, transporte, artes, assuntos ambientais e comunitários e pelos departamentos de *marketing* de grandes empresas.

No contexto da saúde, os interesses são direcionados para aqueles aspectos da QV que podem ser direta ou indiretamente afetados pela saúde dos indivíduos. Portanto, ao tentar-se definir QV relacionada à saúde seria útil iniciar com a definição de saúde. A Organização Mundial da Saúde (OMS) a define como “um estado de completo bem-estar físico, mental e social”. Esta definição tem servido de

base para o desenvolvimento de múltiplas definições de QV, bem como de instrumentos para avaliá-la (THE WHOQOL GROUP, 1995). De uma maneira geral, o termo agrega uma gama de condições e circunstâncias da vida, tais como condições ambientais, inserção social, condições físicas, saúde mental e perspectivas de vida.

Entretanto, também não há um consenso em relação ao significado de “qualidade de vida” na área da saúde. Termos como “estado de saúde”, “estado funcional” e “bem-estar” são utilizados como sinônimos de QV, fazendo seu conceito ficar com limites tênues (FLECK, 2000).

“Estado de saúde” pode ser definido como o nível de saúde de um indivíduo, grupo ou população avaliado de forma subjetiva ou também através de medidas mais objetivas. Considerando “saúde” segundo o conceito mais amplo da OMS, as medidas de estado de saúde podem apresentar pontos em comum com as medidas de QV (FLECK, 2008). Gill & Feinstein (1994), em revisão crítica sobre QV, consideram-na mais do que uma descrição do estado de saúde de um indivíduo: é um reflexo da maneira pela qual um paciente percebe e reage ao seu estado de saúde e a outros aspectos não médicos de sua vida. Estas percepções e reações podem ser mais bem avaliadas se os pacientes puderem expressar o valor que atribuem tanto a sua QV global como a itens individuais que a afetem.

O “estado funcional” pode ser definido como o grau que um indivíduo é capaz de desempenhar seus papéis sociais livres de limitações físicas ou mentais (BOWLING, 1997). Não é considerado um sinônimo de QV, pois um mesmo nível de incapacidade pode coexistir com vários estados existenciais, do desespero à tranquilidade. Ou seja, o estado funcional pode ser um domínio, ou dimensão, da QV.

“Bem-estar” é uma medida que combina a presença de emoções positivas e a ausência de emoções negativas com um senso geral de satisfação com a vida. Apesar da proximidade com o conceito de QV, esta última deve estar inserida no contexto social e cultural tanto do sujeito como do avaliador, não devendo ser reduzida apenas ao equilíbrio entre satisfação e insatisfação (FLECK, 2008).

Outro conceito incluído no termo geral “Qualidade de vida” é o conceito de *Quality Adjusted Life Years* (QALY), que combina sobrevida com QV. É um conceito derivado de teorias econômicas e que considera as preferências dos indivíduos

entre diferentes estados de saúde. O valor quantitativo destas preferências é denominado índice de utilidade (FLECK, 2008). Este conceito será discutido em maiores detalhes na seção 2.1.5.

Mais recentemente, foi introduzido o conceito de *Patient Reported Outcomes* (PRO), que é um termo amplo que inclui qualquer avaliação subjetiva de um indivíduo em relação a elementos de sua saúde, englobando: sintomas, função, bem-estar, percepção sobre os tratamentos, satisfação com os cuidados recebidos, satisfação com a comunicação com os profissionais de saúde e QV (ROTHMAN et al., 2007). Portanto, PRO não é sinônimo de QV, mas engloba este conceito.

A OMS, através do Grupo de Qualidade de Vida definiu QV como “a percepção do indivíduo de sua posição na vida no contexto do sistema cultural e de valores em que ele vive e em relação a seus objetivos, expectativas, padrões e preocupações” (THE WHOQOL GROUP, 1995).

Wenger (1999) citado por Swenson (2004), define QV como um conceito que abrange a maneira pela qual a vida de um indivíduo é afetada por uma doença e por vários componentes do seu tratamento. O constructo de QV deve focar em domínios da vida valorizados pelos pacientes, tais como nível de conforto, senso de bem-estar, habilidade para manter as funções física, emocional e intelectual preservadas e a habilidade para participar de atividades. Qualidade de vida mede a experiência da doença de maneira subjetiva, acrescentando mais informações além das medidas biológicas. Também considera a percepção dos sintomas pelo paciente, a maneira que ele os denomina e os comunica a outras pessoas, a experiência da incapacidade de desempenhar suas atividades normalmente e os esforços feitos para lidar com a doença e ter controle sobre ela.

Vários modelos de compreensão de QV têm sido propostos, embora haja poucos estudos testando sua adequação (FLECK, 2000).

Alguns modelos teóricos propostos foram:

- Modelo da satisfação: Considera características pessoais do indivíduo, condições de vida objetiva em vários domínios e a satisfação com a vida nestes domínios. (FLECK, 2000).
- Modelo combinado importância/satisfação: A satisfação subjetiva e a importância de determinado domínio para o indivíduo são levadas em consideração conjuntamente. Este modelo considera o fato de que um mesmo domínio terá

importância diferente para pessoas diferentes. A partir dele foram criados alguns instrumentos que avaliam QV utilizando um escore ponderal, em que para cada item o paciente atribui um peso para o escore final (FLECK, 2000).

- Modelo do papel ocupacional (*role functioning*): Baseado nas idéias de que satisfação e felicidade relacionam-se com as condições sociais e ambientais requeridas para o preenchimento das necessidades básicas do ser humano. Porém, este modelo parece considerar que estas necessidades sejam universais e estáveis (FLECK, 2000).

- Modelo dinâmico: Considera que a maioria das pessoas possui capacidade para manter o seu nível de satisfação relativamente estável por meio de atividades cognitivas e conativas, mesmo com modificações ambientais. Se um indivíduo está insatisfeito devido a uma discrepância entre suas expectativas e suas condições de vida, pode reduzir a insatisfação modificando as condições ambientais ou as expectativas (FLECK, 2000).

O que há em comum entre a maioria das definições de QV é a ênfase, primeiro, na percepção do paciente, ou seja, a subjetividade e, segundo, na multidimensionalidade do conceito. Enfatizar a percepção do paciente permite diferenciar de “padrão de vida”, que seria uma avaliação objetiva e independente da percepção do indivíduo. Há boas evidências de que o próprio julgamento de uma pessoa sobre a sua saúde tem alto valor preditivo para um desfecho (LEPLEGE & HUNT, 1997).

A multidimensionalidade está representada pela variedade de domínios abordados pelos instrumentos mais abrangentes que permitem avaliar um indivíduo, por exemplo, em seus aspectos físicos, sociais, psicológicos e ambientais (FLECK, 2000).

Na ausência de uma definição operacional unânime, os pesquisadores devem procurar definir qual o conceito que está sendo considerado para seu estudo e identificar quais os domínios serão incluídos e medidos. Como há muitos instrumentos disponíveis, os investigadores devem também justificar a escolha por determinados instrumentos utilizados em sua pesquisa (GILL & FEINSTEIN, 1994).

2.1.2 Instrumentos para medir qualidade de vida

Qualidade de vida é comumente medida através de uma coleção complexa de itens, escalas, domínios e instrumentos (GILL & FEINSTEIN, 1994).

- Um **item** é uma questão única, por exemplo, “Como está seu apetite?”.
- Uma **escala** contém as categorias ou outros mecanismos utilizados para expressar a resposta à questão. Por exemplo, as opções de respostas à pergunta anterior poderiam ser: “Muito bom”, “Bom”, “Moderado”, “Ruim”, “Muito ruim”.
- Um **domínio ou dimensão** identifica um foco particular de atenção como, por exemplo, capacidade funcional ou saúde mental, e pode conter a resposta a um único item ou respostas a vários itens relacionados. Refere-se a uma determinada área de comportamento ou experiência que está sendo medida.
- Um **instrumento** ou **índice** é o conjunto de itens usados para obter os dados desejados. Um instrumento pode conter uma única questão global ou múltiplos itens que podem ou não ser categorizados em domínios separados.

Os resultados gerados por um instrumento podem ser apresentados de duas maneiras: 1) O escore de cada um dos domínios é considerado e são citados individualmente, um após o outro, formando um perfil. 2) Os domínios são agregados para formar um único escore global. Alguns instrumentos permitem realizar ambas as abordagens (GILL & FEINSTEIN, 1994).

Os instrumentos para avaliação de QV podem ser categorizados de acordo com a perspectiva que se propõem a avaliar: os que avaliam QV geral, qualidade de vida ligada à saúde e QV ligada a uma doença específica (FLECK, 2000).

- **Qualidade de vida geral:** Derivada de um referencial social. Abrange de forma ampla os diferentes componentes do constructo QV, fornecendo elementos para compreensão das motivações, desejos, oportunidades e recursos disponíveis para a satisfação e bem-estar de um indivíduo, em diferentes domínios de sua vida. Um exemplo de instrumento desta categoria é o instrumento desenvolvido pela OMS, o WHOQOL.
- **Qualidade de vida ligada à saúde:** Ênfase no estado funcional e senso de bem-estar, porém considera apenas os aspectos diretamente relacionados com a saúde, ou seja, as limitações no funcionamento devidas à doença emocional ou física. Dentro desta categoria, estão todos os instrumentos que enfocam os aspectos

da existência afetados pelo fato de estar doente. Um representante deste grupo, e um dos instrumentos mais utilizados em todo o mundo, é o SF-36.

- Qualidade de vida ligada a uma doença específica: Focaliza aspectos específicos de uma determinada doença em relação à QV. Um exemplo é o *Seattle Angina Questionnaire* (SAQ), criado para avaliação de pacientes que sofrem de angina.

A importância de distinguir QV geral da relacionada à saúde reside no fato de que a primeira abrange fenômenos não médicos como, por exemplo, relações familiares, espiritualidade, satisfação com a vida profissional, aspectos estes que influenciam a QV de um indivíduo independente da presença ou não de uma doença. Se esta distinção for negligenciada, pode-se superestimar o impacto de fatores relacionados à saúde e, inversamente, subestimar o efeito de fenômenos não médicos. Por exemplo, uma paciente com artrite degenerativa leve pode relatar uma QV geral ruim apesar de sua doença estar clinicamente controlada, por conviver com um marido alcoolista agressivo (GILL & FEINSTEIN, 1994).

Gill & Feinstein (1994) recomendam que os pacientes sejam questionados sobre sua QV através das duas maneiras: relatar sua QV global e a QV relacionada à saúde, para assegurar que os efeitos separados dos fatores relacionados à saúde e fatores não médicos sejam determinados e distinguidos.

Muitas vezes os conceitos de “estado de saúde”, “qualidade de vida relacionada à saúde” e “qualidade de vida global” são utilizados como sinônimos, o que gera confusão, pois muitos instrumentos destinados a medir estado de saúde não incluem em seu constructo questões referentes a aspectos mais globais de QV, ou seja, questões não relacionadas à saúde. Considerando que o estado de saúde de um indivíduo pode não ser afetado apenas por uma doença, mas também por fatores como renda, situação profissional e o modo de lidar com a realidade, deve-se ter cuidado ao tentar atribuir uma determinada “qualidade de vida” a um indivíduo sem especificar as dimensões e aspectos que foram levados em consideração (LEPLEGE & HUNT, 1997).

Tem ocorrido alguma confusão entre questionários que são respondidos pelos pacientes e aqueles que refletem as preocupações dos pacientes. Muitas vezes as pessoas são solicitadas a responder questionários que não abordam aquilo que é mais importante para elas. Por exemplo, alguns instrumentos designados para medir

QV em epilepsia tendem a focar o seu conteúdo na frequência e gravidade das convulsões, funcionamento físico e capacidade para o trabalho. No entanto, estudos qualitativos de pessoas com epilepsia têm demonstrado que receberem o rótulo de “epilépticos” e serem socialmente estigmatizados são as principais preocupações dos pacientes (LEPLEGE & HUNT, 1997).

2.1.2.1 Características de um bom instrumento

Dependendo de qual o objetivo de medir QV o instrumento deve ser (GUYATT, FEENY, PATRICK, 1993):

1) Discriminativo: se o objetivo é distinguir pessoas que tem melhor QV daquelas que tem pior QV.

2) Avaliativo: se o foco for a detecção de mudanças na QV através do tempo ou como efeito de um tratamento.

3) Preditivo: se o objetivo é utilizar uma medida alternativa às tradicionais medidas fisiológicas para avaliação de desfechos, a medida de QV pode ser uma variável preditora de desfechos como, por exemplo, taxa de hospitalização, sobrevida, utilização de serviços de saúde e custos.

A seguir serão discutidas as propriedades que os instrumentos devem apresentar para que sejam considerados de bom desempenho (GUYATT, FEENY, PATRICK, 1993).

- Sinal-ruído: Para medidas fisiológicas, reprodutibilidade e acurácia são os atributos necessários para um bom teste. Para instrumentos que medem QV, reprodutibilidade é representada por uma alta razão sinal-ruído e acurácia é representada pela capacidade do instrumento em medir aquilo a que se propôs mensurar. Para instrumentos discriminativos, cujo objetivo é detectar diferenças em escores entre pessoas em um mesmo ponto no tempo, a maneira de quantificar a razão sinal-ruído é chamada **confiabilidade**. Se a variabilidade nos escores entre os pacientes (sinal) é muito maior que a variabilidade no mesmo paciente (ruído), então o instrumento é considerado confiável. Instrumentos confiáveis mostram que pacientes estáveis terão os mesmos resultados depois de repetidas aplicações.

Para instrumentos avaliativos, que se destinam a detectar mudanças nos escores no mesmo indivíduo através do tempo, ou como resultado de um

tratamento, o método de quantificar a razão sinal-ruído é denominado **responsividade**. Representa a capacidade do instrumento para detectar mudanças. A responsividade será diretamente relacionada à magnitude da diferença nos escores em pacientes que melhoraram ou pioraram (sinal) e a variabilidade em escores em pacientes que não mudaram (ruído) (GUYATT, FEENY, PATRICK, 1993).

- Validade:

a) Quando existe um padrão ouro: Embora para instrumentos de avaliação de QV não exista padrão-ouro, algumas situações podem ocorrer nas quais outra medida possa ser considerada referência. Nestas circunstâncias, utiliza-se a **validade de critério** para determinar se um instrumento está medindo o que precisa medir. A medida será válida se os resultados corresponderem àqueles da medida padrão. Exemplo de aplicação da validade de critério é a validação de uma versão mais curta de um instrumento, que seria o teste, como preditora dos resultados da versão mais longa, o padrão ouro.

b) Quando não existe um padrão ouro: Neste caso, a pesquisa em QV tem utilizado estratégias de validação derivadas da psicometria. São elas: **validade de face**: examina se um instrumento parece estar medindo o que pretende medir; **validade de conteúdo**: examina se o domínio de interesse está sendo adequadamente representado pelos itens ou questões, no instrumento; **validade de constructo**: é a mais rigorosa abordagem para estabelecer validade. Um constructo é o embasamento teórico do domínio que queremos medir. Validade de constructo envolve comparações entre medidas e examina as relações lógicas que devem existir entre uma medida e características dos pacientes e grupos de pacientes. A partir de um modelo, ou quadro teórico, que representa a compreensão daquilo que os investigadores estão tentando medir, é possível formular hipóteses sobre o desempenho do instrumento em relação a outras medidas. Os instrumentos são então aplicados a uma população de interesse e a validade será confirmada ou não dependendo da aceitação ou não da hipótese formulada. Por exemplo, um instrumento discriminativo feito para comparar pessoas que receberam ou não tratamento quimioterápico deve ser capaz de distinguir os pacientes nestes dois grupos. Se isto não ocorrer, algo está errado com o desempenho da medida. Outro

exemplo é a validação de um instrumento discriminando grupos de pessoas em relação a aspectos emocionais – os resultados devem correlacionar-se com medidas já validadas de avaliação de função emocional. Os princípios de validação são os mesmos para instrumentos avaliativos, mas a validade é demonstrada quando mudanças nos escores do instrumento correlacionam-se com mudanças em outras medidas relacionadas.

Uma consideração importante é que a validação não termina quando o primeiro estudo com os dados referentes a ela é publicado, mas o processo continua com o uso repetido de um instrumento. Quanto mais for utilizado e o desempenho ocorrer como esperado, maior a confiança em sua validade (GUYATT, FEENY, PATRICK, 1993).

Interpretabilidade é outra propriedade importante para uma medida de QV. Em instrumentos discriminativos devemos saber se o resultado significa ter um prejuízo leve, moderado ou grave na QV; para instrumentos avaliativos, devemos compreender se as mudanças ocorridas nos escores são pequenas, moderadas ou grandes.

A interpretabilidade da maioria das medidas de QV não é evidente por si só. Existem vários métodos disponíveis para compreender a magnitude do efeito na QV. Investigadores podem relacionar mudanças nos escores de um questionário a uma medida funcional bem conhecida, como por exemplo, a classificação funcional do *New York Heart Association*, ou a um diagnóstico clínico, como a alteração no escore necessária para classificar um indivíduo dentro da categoria diagnóstica de depressão, ou o impacto de eventos de vida maiores, como morbidade e mortalidade (GUYATT et al. 1997). Seja qual for a estratégia, é importante que ela seja esclarecida para que os achados tenham significado para uso na prática clínica.

2.1.2.2 Tipos de instrumentos

Os instrumentos para avaliação de QV classificam-se em dois grupos: genéricos e específicos.

Instrumentos genéricos: Podem ser utilizados para comparações entre diferentes tipos de doenças e tratamentos, diversos graus de gravidade de doenças

e diferentes grupos demográficos. Estas medidas são designadas para sumarizar um espectro do conceito de QV que se aplica a muitas condições clínicas e populações diferentes (PATRICK & DEYO, 1989). Este espectro compreende cinco categorias de conceitos: duração da vida, prejuízos, estados funcionais, percepções e oportunidades sociais.

As medidas genéricas são ainda subdivididas em dois grupos: **descritivas ou psicométricas**: geram um “perfil de saúde”; designadas para medir todos os aspectos importantes de QV. O resultado é descrito através de vários escores, um para cada domínio do instrumento, ou um único escore global; **medidas de utilidade**: derivada da teoria de decisão e reflete as preferências dos pacientes por determinados estados de saúde. A vantagem é de possibilitar a combinação de dois desfechos: QV e morte. O resultado é expresso em um único número em uma escala de 0 (morte ou pior estado de saúde possível) a 1 (saúde perfeita). Escore de utilidade reflete a QV e o valor que ela tem para o paciente. Este número associado à taxa de sobrevivência gera a unidade denominada *Quality Adjusted Life Years*(QALY), utilizada como medida de desfecho em análises econômicas de custo-utilidade. Estes instrumentos e seu conceito serão abordados com detalhes na seção 2.1.5.

Instrumentos específicos: Designados para avaliar grupos com um diagnóstico específico ou populações específicas com o objetivo de medir responsividade ou mudanças “cl clinicamente importantes”. Não são necessariamente relacionadas a uma doença, mas também a alguma condição específica como dor lombar, dispnéia, ou a alguma função específica (função sexual, por exemplo) ou população específica, como idosos, crianças doentes (PATRICK & DEYO, 1989).

A tabela 1 apresenta as vantagens e desvantagens de cada grupo.

Tabela 1 - Tipos de instrumentos para avaliação de qualidade de vida

Tipos de instrumentos	Vantagens	Desvantagens
Genéricos		
Descritivos (geram perfis de saúde)	<ul style="list-style-type: none"> • Único instrumento • Detecta efeitos diferenciais em diferentes aspectos de QV • Possibilita comparações entre grupos com diferentes doenças, intervenções 	<ul style="list-style-type: none"> • Pode não focar na área de interesse • Pode não ser responsivo
Medidas de utilidade	<ul style="list-style-type: none"> • Um único número representa o impacto na quantidade e qualidade de vida • Possibilita uso em estudos de custo-utilidade • Incorpora morte como desfecho 	<ul style="list-style-type: none"> • Dificuldade em se determinar valores de utilidade • Não possibilita avaliar o efeito em diferentes aspectos de qualidade de vida • Pode não ser responsivo
Específicos	<ul style="list-style-type: none"> • Clinicamente sensíveis • Podem ser mais responsivos 	<ul style="list-style-type: none"> • Não permite comparações entre grupos com diferentes condições • Pode ser limitada para estudar populações e intervenções

Adaptado de Guyatt, Feeny e Patrick (1993).

A escolha do instrumento adequado depende do objetivo do estudo e do período de avaliação.

Para clínicos envolvidos em ensaios clínicos e com a prática clínica, o interesse pode estar mais voltado para medidas de desfecho designadas a avaliar uma intervenção ou relacionadas a desfechos clínicos. O objetivo principal destes estudos é perceber a mudança ocorrida na QV para o mesmo paciente em dois ou mais pontos no tempo. A capacidade para detectar pequenas mudanças é importante para determinar o poder estatístico de um estudo ou o para o cálculo de tamanho da amostra. Neste caso, as medidas específicas mostram-se mais adequadas.

Para a área da política em saúde, alocação de recursos e epidemiologia, que visam à comparação entre populações, o interesse pode estar focado em avaliar diferenças na QV entre grupos de indivíduos. Comparações entre diferentes doenças, intervenções e subgrupos exigem critérios uniformes para classificar cada população de acordo com sua QV. Portanto, o importante é a detecção de diferenças entre os sujeitos (PATRICK & DEYO, 1989). Com estes objetivos, os

instrumentos genéricos são os mais apropriados.

Patrick & Deyo (1989) preconizam quatro abordagens para uso de instrumentos genéricos e específicos em pesquisa:

a) Instrumento genérico e específico: incluir os dois instrumentos na mesma investigação, embora alguns conceitos possam sobrepor-se entre eles. A vantagem é de poder detectar efeitos em diferentes aspectos de QV, através da medida genérica, sem risco de perder aspectos mais relevantes para uma doença específica.

b) Instrumento genérico e instrumento genérico modificado: algumas medidas genéricas foram modificadas para incluir itens mais específicos à determinada condição. Por exemplo, o *Kidney Disease Quality of Life SF*, por exemplo, inclui todo o SF-36, acrescido de questões relacionadas a aspectos particulares para pacientes com doença renal crônica. Um mesmo estudo poderia comparar a medida genérica com a medida modificada.

c) Instrumento genérico com uma medida suplementar específica: abordagem similar àquela descrita no item a, com a diferença de que a medida específica é construída em uma base conceitual diferente da medida genérica, com mínima sobreposição no constructo. A intenção é capturar aspectos adicionais e específicos para pacientes com determinada condição que não são detectados pela medida genérica.

d) Diferentes instrumentos específicos: incluir, em um mesmo estudo, uma bateria de medidas específicas. Isto também pode ser realizado com medidas genéricas, mas com as específicas tem sido mais comum. Conjunto de instrumentos específicos ou genéricos são frequentemente utilizados em ensaios clínicos e investigações epidemiológicas, quando escalas inteiras, subescalas ou itens individuais dos melhores instrumentos disponíveis são administrados e os efeitos testados para cada medida.

Não há uma única medida que possa ser completa o suficiente para satisfazer todas as necessidades dos investigadores e para populações específicas. Uma das estratégias mais adequadas seria a de utilizar medidas genéricas, padronizadas, com suplementos para situações específicas (PATRICK & DEYO, 1989).

O uso continuado das medidas genéricas torna-se necessário para comparar benefícios de diferentes intervenções em saúde e alocação de recursos. O

conhecimento cumulativo de QV como medida de desfecho utilizando medidas genéricas permitirá conhecer-se o impacto de diferentes doenças e o mérito relativo de diferentes intervenções. Por outro lado, os instrumentos específicos são importantes para identificar preocupações peculiares de pacientes com uma doença específica e avaliar mudanças sutis em seu estado de saúde.

2.1.3 Interpretação dos escores das medidas de qualidade de vida

A representação quantitativa das medidas de QV é geralmente feita através de escores, que podem ser um único número como no caso das medidas de utilidade, ou vários escores medindo diferentes domínios.

A interpretação destes escores nem sempre é compreensível para clínicos e pacientes, principalmente pela falta de um “padrão ouro” para valores de QV que permita comparações. Por exemplo, não existe um limiar para os escores que possa delimitar o que é “normal” ou “disfuncional”, como temos para as medidas biológicas como nível de pressão arterial, nível de hemoglobina ou glicemia. Uma estratégia comumente utilizada por pesquisadores é a comparação de médias de escores entre grupos com diferentes doenças, entre doentes e não doentes e entre dois pontos diferentes no tempo em estudos longitudinais. Porém, mesmo que estas diferenças sejam estatisticamente significativas, algumas questões permanecem não esclarecidas: a) estas diferenças são clinicamente importantes? b) no caso de comparação entre grupos, qual seria o grupo de referência?

A fim de tentar lidar com estes problemas na interpretação de escores de QV, algumas estratégias têm sido utilizadas na literatura (FAYERS & MACHIN, 2007):

- 1) Normas populacionais: São valores de referência dos escores baseados na população geral. Tabelas com valores normativos são obtidas através da aplicação de instrumentos de medida de QV em uma amostra de indivíduos selecionados randomicamente na população geral. Estas tabelas contendo médias de escores em diferentes domínios para subgrupos populacionais como sexo e faixa etária, por exemplo, podem ser utilizadas como “padrão ouro” contra o qual as médias de grupos de pacientes podem ser comparadas. Uma das maneiras de comparar seria através da diferença entre a média populacional (valor esperado) e a média do grupo onde a QV está sendo medida (valor observado). Se os escores normativos

do instrumento utilizado têm um pequeno desvio padrão, então uma pequena diferença entre a média observada e a esperada pode ser considerada importante. Do contrário, se o desvio-padrão populacional é grande, serão necessárias grandes diferenças entre as médias medidas e as médias “padrão” para que a diferença seja relevante. Portanto, diferenças padronizadas, onde a diferença entre médias é dividida pelo desvio padrão pode ser mais fácil de interpretar. Uma das formas de padronização é o cálculo de “tamanho de efeito”, que será discutido mais adiante.

A população escolhida como referência, ou seja, de onde serão gerados os valores normativos, geralmente é a população geral. A maioria dos estudos relatando normatização de instrumentos utilizou desenho transversal e as médias dos escores são apresentadas conforme o sexo e a faixa etária (FAYERS & MACHIN, 2007).

2) Mínima diferença clinicamente importante: É a menor diferença em um escore de QV que seria percebida pelo indivíduo como benéfica e que justificaria mudanças no manejo clínico. Para determinar esta diferença, os pacientes devem ser questionados se observaram uma variação na sua condição e quão importante ela é. A estratégia geralmente consiste em aplicar perguntas padronizadas em pelo menos duas avaliações em períodos diferentes, relacionadas à percepção do indivíduo em relação a mudanças no seu estado. Estas questões são chamadas de “questões de transição” e as opções de resposta podem ser categorizadas, como por exemplo, de “nenhuma mudança” a “muito melhor (ou pior)”. A maioria dos pesquisadores tem utilizado pelo menos sete categorias. Estas respostas são depois comparadas aos escores obtidos no instrumento de QV medido concomitantemente (FAYERS & MACHIN, 2007).

3) Tamanho de efeito: Quando não há a informação de normas populacionais ou diferença mínima clinicamente importante disponível para comparações, o cálculo do tamanho de efeito pode ser útil. Cohen (1988) propôs esta padronização a fim de simplificar a estimativa para cálculo de tamanho de amostra. Baseado em suas experiências nas ciências sociais, ele sugeriu a seguinte classificação para o tamanho de efeito: índices de 0,2 a 0,5 seriam efeitos “pequenos”, de 0,5 a 0,8 seriam “moderados” e acima de 0,8 “grandes”. Estes valores arbitrários têm se mostrado aplicáveis até hoje em vários campos de pesquisa. Na área de QV, duas formas de calcular o tamanho de efeito são utilizadas:

a) Média de resposta padronizada (*Standardized Response Mean* - SRM) ou tamanho de efeito de Cohen: Representa a mudança na média dos escores da QV medida em um mesmo indivíduo em dois períodos diferentes, dividida pelo desvio-padrão (DP) desta mudança nos escores. A fórmula é:

$$\text{SRM} = \frac{\text{Média}_{\text{tempo1}} - \text{Média}_{\text{tempo2}}}{\text{DP}_{\text{diferença}}}$$

É a medida mais amplamente utilizada. O $\text{DP}_{\text{diferença}}$ deve ser estimado de pacientes com QV mais estável. Este dado pode estar disponível dos estudos de confiabilidade teste-reteste realizados no desenvolvimento do instrumento, uma vez que estes estudos geralmente são conduzidos em populações não doentes. Quando esta informação não estiver presente as diferenças de desvio padrão podem ser obtidas da própria população do estudo que está sendo realizado. Esta fórmula também pode ser utilizada para comparar médias entre diferentes grupos independentes. Neste caso, substituem-se as médias no tempo 1 e tempo 2 na equação acima por médias do grupo 1 e grupo 2 e o desvio padrão da diferença é substituído pelo desvio-padrão agregado (combinação dos desvios-padrão de cada grupo).

b) Tamanho de efeito “estatístico” (*effect size* - ES): É a diferença na média dos escores entre dois períodos diferentes, dividida pelo desvio-padrão do escore no tempo 1.

$$\text{ES} = \frac{\text{Média}_{\text{tempo1}} - \text{Média}_{\text{tempo2}}}{\text{DP}_{\text{tempo1}}}$$

Esta medida também pode ser utilizada para comparar dois grupos independentes. Para isto, o DP de um dos grupos, geralmente o grupo controle, substitui o $\text{DP}_{\text{tempo1}}$ na fórmula acima. Esta versão do tamanho de efeito é conhecida como delta de Glass (FAYERS & MACHIN, 2007).

2.1.4 Instrumentos genéricos de avaliação de QV: SF-36 e WHOQOL-breve

Existem diversos instrumentos para avaliação de QV genéricos disponíveis na literatura. Dois deles foram utilizados no projeto que deu origem a presente tese, com o objetivo de prover escores normativos para uma amostra da população da cidade de Porto Alegre. São eles o SF-36 e o WHOQOL-breve, que serão descritos nas próximas seções.

2.1.4.1 *Medical Outcomes Study Short-Form 36 (SF-36)*

O SF-36 foi criado a partir da necessidade de ter-se um instrumento padronizado que abordasse conceitos de saúde geral, não específico para qualquer condição clínica e que fosse compreensível, de fácil aplicação e com boas propriedades psicométricas. A base conceitual para o desenvolvimento do SF-36 foram os conceitos de “*status* funcional” e “bem-estar” descritos em definições aceitas de “estado de saúde” (WARE JR. & SHERBOURNE, 1992). Desta forma, o conceito de QV considerado na construção deste instrumento foi o de QV relacionada à saúde, enfatizando o impacto específico que a prevenção e o tratamento de uma doença têm no “valor de estar vivo”.

Os atributos escolhidos para desenvolver o SF-36 foram: funcionamento físico, social, papel ocupacional (*role-functioning*), percepções da saúde geral, saúde mental, dor e vitalidade.

O SF-36 é uma medida multidimensional, com 36 questões. É dividida em oito domínios. É um instrumento genérico, sendo útil para comparar grupos tanto na população geral quanto em populações com doenças específicas, comparar o impacto relativo das doenças, avaliar os benefícios produzidos por diversos tratamentos e como rastreamento para pacientes individuais (WARE JR., 2000).

A utilidade do SF-36 em estimar o impacto das doenças está representada em artigos descrevendo escores do instrumento em vários grupos de pacientes totalizando mais de 130 condições clínicas. Entre as situações mais frequentemente estudadas estão artrite, dor lombar, depressão, diabetes e hipertensão (WARE JR., 2000).

Taxonomia do SF-36

A tabela 2 descreve a estrutura do SF-36.

A taxonomia tem três níveis: 1) itens; 2) escalas que agregam dois a 10 itens cada, 3) duas medidas sumarizadas (componentes) que agregam os domínios.

Todos os itens do SF-36 são usados para pontuar os oito domínios, exceto o item 2, que se refere a um auto-relato de transição de saúde. Cada item faz parte de somente um domínio. Após recalibrar dois itens e inverter o escore de nove itens, as respostas aos itens são somadas. Escores mais altos representam melhores estados de saúde. Um escore para *missing* é computado se itens de uma escala não foram respondidos. A escala dos escores vai de 0 a 100, 0 indicando o estado de saúde menos favorável, 100 o mais favorável e os escores entre eles representam o percentual do possível escore total alcançado (WARE JR., 2000).

Através de análise fatorial, cada domínio foi correlacionado com componentes de saúde física e mental para formar as duas medidas sumarizadas, física e mental. Os domínios capacidade funcional, aspectos físicos e dor têm melhor correlação com o componente físico e contribuem para a maior parte do escore da medida sumarizada física. O componente mental correlaciona-se melhor com os domínios saúde mental, aspectos emocionais e aspectos sociais, contribuindo também em grande parte para a medida sumarizada mental.

Os três domínios vitalidade, estado geral de saúde e aspectos sociais têm correlações com ambos componentes, mental e físico.

Tabela 2 - Taxonomia do SF-36

Itens	Domínios	Medidas sumarizadas
3a. Atividades vigorosas 3b. Atividades moderadas 3c. Levantar ou carregar mantimentos 3d. Subir vários lances de escada 3e. Subir um lance de escada 3f. Curvar-se, ajoelhar-se ou dobrar-se 3g. Andar mais de um quilômetro 3h. Andar vários quarteirões 3i. Andar um quarteirão 3j. Tomar banho ou vestir-se	Capacidade Funcional	Componente físico
4a. Diminuir a quantidade de tempo 4b. Realizar menos tarefas 4c. Limitação em atividades 4d. Dificuldade no trabalho	Aspectos físicos	
7. Magnitude da dor 8. Interferência da dor	Dor	
1. Avaliação global da saúde 11a. Adoecer mais facilmente 11b. Tão saudável quanto 11c. Saúde vai piorar 11d. Saúde excelente	Estado geral de Saúde	Componente mental
9a. Vigor/vontade/força 9e. Energia 9g. Esgotamento 9i. Cansaço	Vitalidade	
6. Interferência na vida social 10. Interferência no tempo da vida social	Aspectos sociais	
5a. Diminuir quantidade de tempo 5b. Realizar menos tarefas 5c. Cuidado com atividades	Aspectos emocionais	
9b. Pessoa nervosa 9c. Deprimido 9d. Calmo/Tranquilo 9f. Desanimado/abatido 9h. Feliz	Saúde mental	

Adaptado de Ware Jr. (2000).

A confiabilidade dos oito domínios e das duas medidas sumarizadas tem sido estimada usando os métodos de consistência interna e teste-reteste. As publicações relatando estatística de confiabilidade do SF-36 têm demonstrado, com raras exceções, índices acima de 0,70, e até 0,80, para os diferentes domínios e, para as medidas sumarizadas, os coeficientes têm excedido 0,90. Revisão dos primeiros 15 estudos publicados revelou que a média dos coeficientes de confiabilidade para cada um dos oito domínios foi igual ou maior que 0,80, exceto para o domínio aspectos sociais, com média 0,76 (WARE JR., 2000).

A validade de conteúdo do SF-36 tem sido comparada com aquela de outros instrumentos amplamente utilizados. Comparações sistemáticas indicam que o SF-36 inclui oito dos conceitos de saúde medidos com maior frequência. As áreas que são comumente avaliadas, mas não estão incluídas no SF-36 são: adequação do sono, funcionamento cognitivo, preocupações com a saúde, funcionamento familiar, auto-estima, alimentação, recreação e lazer, comunicação, função sexual.

O SF-36 é um questionário que pode ser administrado por: auto-aplicação, administração por computador, por um entrevistador treinado ou por telefone e é adequado para pessoas acima dos 14 anos de idade. Pode ser administrado em 5 a 10 minutos com alto grau de aceitabilidade e qualidade dos dados (WARE JR., 2000).

A tabela 3 apresenta um sumário do conteúdo do SF-36 com as médias dos escores em cada domínio para a população dos Estados Unidos, país de origem do instrumento, demonstrando também a confiabilidade de cada escala e o significado do menor e do maior escore.

Tabela 3 - Sumário do conteúdo do SF-36

Domínios	Nº de itens	Média	DP	Confiabilidade	Menor escore possível (<i>floor</i>)	Maior escore possível (<i>ceiling</i>)
Capacidade funcional	10	84,2	23,3	0,93	Muito limitado em realizar todas as atividades físicas, incluindo banhar-se e vestir-se	Realiza todos os tipos de atividades físicas incluindo as mais vigorosas sem limitações devidas a saúde
Aspectos físicos	4	80,9	34,0	0,89	Problemas com o trabalho ou outras atividades diárias como consequência da saúde física	Nenhum problema com trabalho ou outras atividades diárias
Dor	2	75,2	23,7	0,90	Dor muito severa e extremamente limitante	Nenhuma dor e nenhuma limitação devido à dor
Estado geral de saúde	5	71,9	20,3	0,81	Avalia sua saúde geral como muito ruim e acredita que ela piorará	Avalia sua saúde pessoal como excelente
Vitalidade	4	60,9	20,9	0,86	Sente-se cansado ou esgotado todo o tempo	Sente-se cheio de energia e vigor todo o tempo
Aspectos sociais	2	83,3	22,7	0,68	Interferência extrema e frequente de problemas físicos e emocionais nas atividades sociais	Atividades sociais não sofrem interferência por problemas físicos ou emocionais
Aspectos emocionais	3	81,3	33,0	0,82	Problemas com o trabalho ou outras atividades diárias como consequência de problemas emocionais	Nenhum problema com trabalho ou atividades diárias.
Saúde mental	5	74,7	18,1	0,84	Sentir-se nervoso ou deprimido todo o tempo	Sentir-se feliz, calmo e tranquilo todo o tempo

DP = Desvio padrão

Adaptado de Ware Jr. (2000).

Os oito domínios são ordenados de acordo com seu construto. A validade de cada um deles foi avaliada através de estudos de análise fatorial. O primeiro é o de capacidade funcional, que tem sido a melhor medida de saúde física; o último domínio, saúde mental, foi o que teve o melhor desempenho nos testes de validação para avaliar saúde mental em diversos estudos. Contudo, o domínio saúde mental é a medida com pior desempenho para avaliar componente físico e o domínio capacidade funcional é a pior para avaliação de componente mental. Os domínios de vitalidade e estado geral de saúde são válidas para avaliação de ambos componentes, físico e mental (WARE JR., 2000).

O SF-36 tem sido tão amplamente aplicado devido a sua brevidade e fácil compreensão. Estudos descritivos realizados tanto na população geral como em grupos com diagnósticos específicos e ensaios clínicos demonstram que o SF-36 é muito útil para avaliar diferenças entre doentes e não doentes e para estimar o impacto relativo de diferentes condições médicas. Experiência em estudos longitudinais sugere que esta medida é também uma ferramenta de grande utilidade para avaliar os benefícios de diferentes tratamentos (WARE JR., 2000).

Há evidências de que o SF-36 é um instrumento apropriado para uso em diversos grupos de pacientes portadores tanto de doenças clínicas como psiquiátricas (McHORNEY et al., 1994).

O SF-36 foi traduzido para a língua portuguesa e validado no Brasil por Ciconelli et al. (1999). O estudo foi realizado com uma população de pacientes com artrite reumatóide, utilizando protocolo desenvolvido de acordo com etapas propostas pelos coordenadores do *International Quality of Life Assessment Project* (IQOLA), projeto que reúne vários países na tarefa de traduzir e validar o SF-36 em vários idiomas e culturas, adotando normas padronizadas (WARE JR. et al., 1995). A tabela 4 mostra média e desvio-padrão dos escores de cada domínio para a população estudada e a reprodutibilidade intra e interobservador.

Tabela 4 - Valores obtidos e reprodutibilidade para cada domínio do questionário SF-36 e o intervalo de tempo entre as duas aplicações na validação para população brasileira com artrite reumatóide

Domínios	Média	Desvio-padrão	Coeficiente de correlação de Pearson	
			Intra-observador	Inter-observador
Capacidade funcional	66,50	25,95	0,8044**	0,8101**
Aspectos físicos	59,50	44,28	0,6392**	0,6271**
Dor	63,96	21,26	0,5426**	0,5542**
Estado geral de saúde	68,22	22,89	0,8468**	0,6186**
Vitalidade	66,30	26,24	0,6523**	0,7828**
Aspectos sociais	87,00	19,39	0,7595**	0,7158**
Aspectos emocionais	66,66	42,05	0,4426*	0,6884**
Saúde Mental	70,32	25,70	0,6927**	0,8041**
Intervalo de tempo entre as duas aplicações (dias)	9,00	3,00		

* $p < 0,01$ ** $p < 0,001$
Ciconelli et al., 1999.

Atualmente o SF-36 é o instrumento para medida de QV mais amplamente utilizado no Brasil. No entanto, não há dados normativos para a população geral disponíveis em nenhuma das regiões do país, dificultando a interpretação dos escores obtidos em diferentes grupos de pacientes.

2.1.4.2 World Health Organization Quality of Life Instrument-bref (WHOQOL-breve)

O Grupo de Qualidade de Vida da OMS desenvolveu um instrumento para avaliação de QV levando em consideração a necessidade de uma medida criada a partir de uma perspectiva transcultural para uso internacional (WHOQOL-breve, 1998a). Vários centros culturalmente diversos foram envolvidos em operacionalizar os domínios de QV, seleção de itens, confecção das escalas de resposta e testagem do instrumento.

O primeiro passo foi a definição do conceito de QV no qual o construto do instrumento seria baseado. Qualidade de vida foi então definido pela OMS como “a percepção do indivíduo de sua posição na vida no contexto de sua cultura e sistemas de valores nos quais ele vive e em relação aos seus objetivos, expectativas, padrões e preocupações”. É um conceito amplo, abrangendo a saúde física de um indivíduo, seu estado psicológico, nível de independência, relações sociais, crenças pessoais e sua relação com o ambiente. Esta definição enfatiza a visão de que o conceito de QV é subjetivo e inclui aspectos positivos (por exemplo: papel na sociedade, nível de satisfação e mobilidade) e aspectos negativos (p.ex. dependência de medicação, dor, fadiga) da vida e é multidimensional (THE WHOQOL GROUP, 1995).

Inicialmente foi desenvolvido um instrumento com 100 questões (WHOQOL-100, 1998b), organizado em seis domínios: físico, psicológico, nível de independência, relações sociais, ambiental e espiritualidade/religiosidade/crenças pessoais. Cada domínio é formado por vários subdomínios (facetas). Contudo, a necessidade de instrumentos curtos que demandem pouco tempo para preenchimento e com boas características psicométricas fez com que o Grupo de Qualidade de Vida da OMS desenvolvesse uma versão abreviada do WHOQOL-100, o WHOQOL-breve (1998a).

O WHOQOL-breve (1998a) é composto por 26 questões, duas de QV geral e as demais representam cada uma das 24 facetas que compõem o instrumento original. Diferente do WHOQOL-100 em que cada uma das 24 facetas é avaliada a partir de quatro questões, na versão abreviada cada faceta é avaliada por apenas uma questão. Os dados que originaram esta versão foram extraídos do teste de campo de 20 centros em 18 países diferentes. O instrumento é composto por quatro domínios: físico, psicológico, relações sociais e meio ambiente. A tabela 5 descreve os domínios e facetas do WHOQOL-breve.

O WHOQOL-breve foi traduzido em vários idiomas e validado em diversos países. No Brasil, este trabalho foi realizado por Fleck et al. (2000) na Universidade Federal do Rio Grande do Sul. Em uma primeira etapa foi feita a seleção de itens e a análise confirmatória da estrutura do instrumento utilizando dados internacionais e depois cada centro conduziu análises com dados locais de cada país. A aplicação do teste de campo da versão em português foi realizada em uma amostra de 300 indivíduos, sendo 50 voluntários-controles e 250 pacientes do Hospital de Clínicas

de Porto Alegre, 125 internados e 125 ambulatoriais, das áreas de clínica médica, cirurgia, psiquiatria e ginecologia.

Tabela 5 - Domínios e facetas do WHOQOL-breve

Domínio 1 – Domínio físico	
1.	Dor e desconforto
2.	Energia e fadiga
3.	Sono e repouso
9.	Mobilidade
10.	Atividades da vida cotidiana
11.	Dependência de medicação ou de tratamentos
12.	Capacidade de trabalho
Domínio 2 – Domínio psicológico	
4.	Sentimentos positivos
5.	Pensar, aprender, memória e concentração
6.	Auto-estima
7.	Imagem corporal e aparência
8.	Sentimentos negativos
24.	Espiritualidade/religião/crenças pessoais
Domínio 3 – Relações sociais	
13.	Relações pessoais
14.	Suporte (Apoio) social
15.	Atividade sexual
Domínio 4 – Meio Ambiente	
16.	Segurança física e proteção
17.	Ambiente no lar
18.	Recursos financeiros
19.	Cuidados de saúde e sociais: disponibilidade e qualidade
20.	Oportunidades de adquirir novas informações e habilidades
21.	Participação em, e oportunidades de recreação/lazer
22.	Ambiente físico (poluição/ruído/trânsito/clima)
23.	Transporte

Fleck et al., 2000.

O WHOQOL-breve produz um perfil de QV. Além dos escores de cada um dos quatro domínios, há também dois itens que são examinados separadamente: a questão 1 pergunta sobre a percepção do indivíduo sobre sua QV em geral e a questão 2 sobre percepção de sua saúde geral.

O escore médio dos itens de cada domínio é utilizado para calcular o escore do domínio. As médias dos escores são então multiplicadas por 4 a fim de possibilitar a comparação com os escores do WHOQOL-100.

A versão em português do WHOQOL-breve apresentou boa consistência interna, validade discriminante, validade concorrente, validade de conteúdo e confiabilidade teste-reteste (FLECK et al., 2000). A tabela 6 demonstra a avaliação da consistência interna, as médias dos escores de pacientes e controles, com a análise de variância e a fidedignidade teste-reteste através dos coeficientes de correlação.

Tabela 6 - Análise de variância do domínio em relação a normal X paciente, coeficiente de fidedignidade de Cronbach dos domínios e questões e coeficiente de correlação dos escores dos diferentes domínios entre teste e reteste na aplicação da versão em português do WHOQOL-breve

Itens	Médias (desvio-padrão)		p	Coeficiente de Cronbach	Coeficiente de correlação teste-reteste
	Controle	Paciente			
Domínios 26 questões				0,77 0,91	
Domínio 1	66,4 (8,4)	53,6 (12,8)	0,0001	0,84	0,81
Domínio 2	62,4 (8,4)	58,0 (11,2)	0,01	0,79	0,69
Domínio 3	62,0 (10,4)	61,2 (14,4)	0,66	0,69	0,80
Domínio 4	56,0 (8,4)	53,6 (9,6)	0,06	0,71	0,75

Fleck et al., 2000.

O WHOQOL-breve é um instrumento genérico, auto-aplicável, possibilitando também a administração por entrevistador. Pode ser utilizado em ensaios clínicos para estabelecer escores basais em várias áreas e observar mudanças na QV após intervenções. É útil em situações clínicas onde não há possibilidade de cura de uma doença, apenas recuperação ou remissão parciais e o tratamento é apenas paliativo.

Para pesquisas epidemiológicas, o instrumento permite detalhados dados sobre QV de uma população em particular, facilitando a compreensão das doenças e o desenvolvimento de métodos de tratamento. E na prática clínica, a avaliação através do WHOQOL possibilita ao médico reconhecer quais áreas da vida do paciente estão sendo mais afetadas pela doença ou tratamento, auxiliando na tomada de decisões.

Estudos relatando dados normativos do WHOQOL-breve para população geral são escassos mesmo internacionalmente (NOERHOLM et al., 2004; OHAERI, AWADALLA, GADO, 2009) e não estão disponíveis no Brasil.

2.1.5 Medidas de utilidade

Nos últimos anos tem aumentado o interesse em incorporar as medidas de QV em análises econômicas em saúde como uma medida de desfecho para avaliar as diferentes intervenções e programas.

Contudo, os instrumentos de QV psicométricos descritos acima, do ponto de vista dos economistas, não são adequados para serem incorporadas como uma medida de desfecho em estudos de custo-efetividade pelos seguintes aspectos: 1) os instrumentos geralmente não produzem um único escore de QV, mas sim um perfil de escores em diferentes domínios. Não é possível comparar uma melhora em um domínio com outro domínio, por exemplo, o domínio físico não pode ser comparado ao psicológico. Para as análises econômicas é necessário um único número que represente uma mudança ocorrida com uma intervenção e que possibilite comparações entre diferentes intervenções que levam a desfechos diversos; 2) os escores dos instrumentos não são baseados nas preferências dos indivíduos pelos vários desfechos possíveis e não significa que escores mais altos representem mais preferência das pessoas por aquele estado de saúde; 3) pelo fato de que os escores dos instrumentos psicométricos não são calibrados em uma escala entre 0 e 1 onde 0 representa morte e 1 representa saúde perfeita, eles não podem ser utilizados para combinar QV com sobrevida para gerar o QALY, um índice de uso comum em estudos de custo-efetividade (DRUMMOND et al., 2005).

A fim de inserir a QV como um dos desfechos utilizados em análises econômicas, economistas e pesquisadores envolvidos com medida de QV

desenvolveram instrumentos adequados a este propósito, gerando as chamadas medidas de utilidade. Mais recentemente, as medidas de utilidade foram combinadas às medidas psicométricas, dando origem às medidas de QV baseadas em preferências (*preference-based measures*).

2.1.5.1 Base teórica

A teoria de *utilidade* foi desenvolvida no campo da economia como uma maneira de quantificar e analisar a maneira pela qual as pessoas fazem escolhas.

Em 1944, John Von Neumann, matemático e Oscar Morgenstern, economista, publicaram sua teoria de tomada de decisão racional sob condições de incerteza, a chamada “Teoria de Utilidade de Von Neumann e Morgenstern” (Von NEUMANN & MORGENSTERN, 1944). Estes autores desenvolveram um modelo normativo em relação ao comportamento das pessoas ao tomar decisões, ou seja, como indivíduos racionais deveriam fazer escolhas quando confrontados com situações de incerteza (DRUMMOND et al., 2005).

Devido ao fato de esta teoria se adequar a qualquer bem quantificável, seu uso também foi disseminado naturalmente para a área da saúde, os chamados “índices de utilidade de estados de saúde” (TORRANCE, 1976).

A abordagem de Von Neumann & Morgenstern (1944) consiste em derivar uma função de utilidade para um indivíduo baseada nas preferências deste indivíduo entre pares de “apostas” (*gamble*). Estas “apostas” consistem em escolhas entre dois cenários com probabilidades de ocorrência diferentes. Para que esta função de utilidade seja capaz de representar preferências algumas suposições são necessárias. Essas suposições são geralmente denominadas na literatura como os “axiomas da teoria de utilidade de Von Neumann & Morgenstern”. Considerando uma situação onde há desfechos mutuamente exclusivos, por exemplo, A_1, A_2, \dots, A_n , estas suposições seriam as seguintes: a) para cada par de desfechos, o indivíduo prefere um dos dois ou é indiferente entre eles; b) as preferências são transitivas, que significa que se A_1 é preferido a A_2 e A_2 é preferido a A_3 , então A_1 é preferido a A_3 ; c) pressuposto de independência: apostas compostas, ou seja, uma aposta na qual a alternativa é outra aposta, pode ser equivalente a duas apostas simples se as suas probabilidades são equivalentes. Por exemplo,

considerando duas apostas y e y' onde y é formada pelo desfecho x_1 com probabilidade p_1 e o desfecho x_2 com probabilidade $(1 - p_1)$, indicada simbolicamente por $y = \{ p_1, x_1, x_2 \}$ e $y' = \{ p_2, x_1, x_2 \}$. Este axioma implica que um indivíduo seria indiferente entre a aposta composta (p, y, y') e a aposta simples probabilisticamente equivalente $\{ p p_1 + (1 - p) p_2, x_1, x_2 \}$; d) pressuposto de continuidade: se há 3 desfechos tais que x_1 é preferido a x_2 , que é preferido a x_3 , há alguma probabilidade p na qual o indivíduo é indiferente entre o desfecho x_2 com probabilidade de 100% (certeza) ou uma aposta formada pelo desfecho x_1 com probabilidade p e o desfecho x_3 com probabilidade $1 - p$. Estas suposições formam a base teórica da técnica *standard gamble* (SG), um método para medir preferências (FEENY & TORRANCE, 1989; DRUMMOND et al., 2005).

A função de utilidade de Von Neumann e Morgenstern é também conhecida como a “teoria da utilidade esperada” (*expected-utility theory*): a utilidade de uma aposta é igual à utilidade esperada dos seus desfechos componentes (FEENY & TORRANCE, 1989).

É importante ressaltar que a Teoria de Utilidade de Von Neumann e Morgenstern e suas medidas de preferência associadas, os chamados índices de utilidade (*utilities*) não possuem nenhuma relação com “utilitarismo” e não têm o mesmo significado de “utilidade” na linguagem normal, no sentido de “ser útil”. Neste caso, utilidade representa, como descrito acima, o valor da preferência do indivíduo por um determinado desfecho quando esta escolha é feita em condições de incerteza.

Os termos “utilidade”, “valor” e “preferência” são muitas vezes usados como sinônimos, mas existem diferenças entre eles. Preferência é um termo “guarda-chuva” que descreve um conceito mais geral e utilidade e valor são tipos diferentes de preferências, dependendo de como a medida é realizada. Dois aspectos da medida são essenciais para classificar o tipo de preferência: se os desfechos são certos ou incertos e se o indivíduo deve fazer uma avaliação (*rating*) de cada desfecho ou fazer uma escolha entre eles. A tabela 7 descreve a classificação das preferências conforme a estratégia de medida (DRUMMOND et al., 2005).

Tabela 7 - Métodos para medir preferências

Tipo de resposta	Forma de elaboração da questão	
	Certeza (valores)	Incerteza (utilidade)
Escalonar	Escalas de avaliação Escalas categóricas Escala análogo visual	
Escolha	<i>Time trade-off</i> (equivalência temporal) Comparações pareadas Equivalência	<i>Standard gamble</i> (jogo padrão)

Adaptado de Drummond (2005).

Uma questão elaborada sob condições de certeza exige que o indivíduo compare dois ou mais desfechos e escolha entre eles ou classifique-os em uma escala, assumindo que os desfechos ocorrerão com certeza. As preferências geradas por este tipo de questionário não são consideradas conceitualmente índices de utilidade, mas sim valores. Por outro lado, uma questão formulada em uma situação de incerteza confronta o indivíduo com uma escolha onde pelo menos uma das alternativas é incerta, ou seja, contém probabilidades associadas a ela (DRUMMOND et al., 2005). O valor das preferências obtidos através desta técnica são os chamados índices de utilidade. Basicamente, a diferença entre as formas de questionamento é que o método de certeza não captura o comportamento do indivíduo em relação a risco, enquanto o método baseado em incerteza o faz. Atitude em relação ao risco é um termo bastante utilizado em medidas de preferência. Se uma pessoa evita situações de maior risco e busca alternativas menos arriscadas, ela é considerada avessa ao risco; indiferença entre duas situações representa neutralidade em relação a risco e preferência por uma “aposta”, ou seja, por desfechos incertos, indica uma atitude de busca ao risco (*risk-seeking*) (DRUMMOND et al., 2005).

Extrapolando o conceito derivado da teoria econômica para as medidas de QV, o termo utilidade refere-se à preferência dos indivíduos por um desfecho, que neste caso é um estado de saúde. As medidas de utilidade avaliam o valor ou conveniência de um determinado estado de saúde em uma escala de valores.

Os valores de preferências podem ser ordinais ou cardinais. Os valores ordinais representam preferências por desfechos expressos através de um ordenamento do desfecho mais preferido para o menos preferido. Para valores cardinais, um número deve ser associado ao desfecho de forma que represente a força da preferência por aquele desfecho comparado aos outros. Estes números devem ser medidos através de uma escala com as seguintes características: a) deve ser intervalar, isto é, semelhante à escala de temperatura em Fahrenheit, por exemplo, onde os pontos podem receber valores arbitrários, sem um zero natural; b) deve ter intervalos iguais de forma que uma mudança de, por exemplo, 0,2 para 0,3 tenha o mesmo significado de uma mudança de 0,8 para 0,9 (DRUMMOND et al., 2005).

As propriedades das funções de utilidade descritas conforme os axiomas da teoria de Von Neumann e Morgenstern não são suficientes para fazer comparações válidas de preferências entre indivíduos, uma vez que elas caracterizam um comportamento individual para tomada de decisões e os valores das preferências poderiam ser infinitos. Uma das soluções encontradas para lidar com este problema foi desenvolver pontos de referência para desfechos em saúde para as análises dos índices de utilidade. Ficou estabelecido que os valores de utilidade fossem comparados dentro de uma escala entre 0 e 1 onde 0 representaria morte e 1 um estado de saúde perfeita (FEENY & TORRANCE, 1989).

O principal uso dos índices de utilidade é guiar decisões sob incerteza ou sob escassez de recursos. Aqui, os valores servem como “fatores de ajustamento para qualidade” para o cálculo dos QALYs, uma medida de desfecho preconizada por diretrizes internacionais (WEINSTEIN et al., 1996) nas análises de custo-efetividade. Podem também ser usados como desfechos em ensaios clínicos e para auxiliar os pacientes a refletir sobre o que realmente importa para eles (TSEVAT, 2000). A aplicação prática dos valores de preferências será vista com mais detalhes na seção 2.2.

2.1.5.2 Desenvolvimento das medidas

Os vários métodos para medida de preferências em geral incluem três processos: 1) Definir um grupo de estados de saúde de interesse, 2) identificar

avaliadores que possam julgar o valor de cada estado; 3) agregar todos os valores obtidos dos avaliadores e chegar a um valor médio para cada estado de saúde. A partir desta regra geral, algumas questões devem ser consideradas para a medida de preferências e serão descritas nos tópicos abaixo (FROBERG & KANE, 1989a).

Quais são os domínios de saúde relevantes?

Ao desenvolver um instrumento para medida de preferências, o pesquisador deve reconhecer quais são os domínios importantes para a população estudada, que podem abranger desde aspectos de saúde até meio-ambiente, suporte social, religiosidade, etc. Como regra, recomenda-se não mais que nove atributos (domínios), uma vez que estudos têm demonstrado que os seres humanos podem processar simultaneamente apenas cinco a nove “porções” de informação. Para cada domínio, um número de níveis é definido representando incrementos graduais de funcionamento, de bom a ruim. A descrição de cada nível geralmente foca na função mais do que no diagnóstico clínico. Os estados de saúde são geralmente formados através da seleção de um nível de cada domínio. Cada estado de saúde tem um valor potencial associado e ele, valor este então chamado de “preferência” ou “utilidade”, sendo um valor cardinal ou índice. A fim de obter estes valores o investigador necessita organizar um protocolo para medi-los. Este processo para obtenção dos valores das preferências será discutido nos itens a seguir.

Qual a estratégia mais adequada para a agregação dos valores obtidos na medida de preferências por vários estados de saúde?

Duas abordagens têm sido aplicadas para a construção do sistema matemático utilizado para calcular os índices de utilidade após obter medidas das preferências por diferentes estados de saúde.

Abordagem holística

Esta abordagem exige que o avaliador atribua valores para cada possível estado de saúde entre vários estados formados por uma combinação de domínios. Tal abordagem foi a estratégia mais usada inicialmente. O respondente deve avaliar

cada estado de saúde com múltiplos domínios (estados denominados na literatura de “multiatributos”), mas os efeitos separados de cada domínio não são analisados. Um exemplo desta abordagem é a utilizada por Patrick (PATRICK, BUSH, CHEN, 1973), que definiu 29 níveis de determinadas funções de um indivíduo, cinco grupos de idade e 42 tipos de sintomas. Os níveis das funções formaram um conjunto contendo três domínios: atividade física, mobilidade e atividade social. Os níveis das funções, os grupos de idade e os sintomas foram combinados de uma forma a gerar uma matriz que descreve um universo de condições que poderiam estar presentes em um indivíduo. As diferentes combinações formaram 400 estados de saúde (figura 1). Por exemplo:

6-17 anos Caminha livremente Viaja livremente Não realiza atividades maiores, mas realiza auto-cuidado Tem tosse, chiado no peito e falta de ar

Fig. 1 - Exemplo de um estado de saúde “multiatributo”.

O estado de saúde foi apresentado como uma vinheta para que os avaliadores julgassem qual sua preferência por diferentes estados, tendo de colocá-los em uma escala de 1 a 16. Patrick, Bush e Chen (1973) então calcularam a média de todos os valores dados pelos avaliadores para cada estado e esta média seria o valor daquele estado de saúde na escala. As maiores limitações desta abordagem é que não é possível avaliar qual a importância dada a cada domínio no estado de saúde, uma vez que ele é avaliado como um todo. Por exemplo, para o avaliador, o domínio “idade” é mais importante do que “poder viajar livremente”? Outro problema é o impacto causado no avaliador por ter de julgar uma quantidade grande de estados de saúde, sobrecarregando-o de informações. Por este motivo, mais recentemente a abordagem holística tem sido substituída pela decomposta.

Abordagem “decomposta”

Permite que o investigador obtenha valores para todos os estados de saúde sem exigir que cada estado seja avaliado. Simplifica a tarefa de avaliação porque

expressa o valor global de um estado de saúde como uma função decomposta do valor de cada domínio, reduzindo o número de avaliações necessárias para construir a escala de valores para o grupo completo de estados de saúde. Dentro da categoria geral da abordagem decomposta, pode-se distinguir entre: 1) procedimentos que separam o processo de avaliação em tarefas menores (modelos explicitamente decompostos); 2) procedimentos que tentam desenvolver um modelo algébrico a partir das preferências de um grupo de estados de saúde “multiatributos” (modelos estatisticamente inferidos);

O modelo algébrico, semelhante à abordagem holística, exige que o avaliador julgue estados de saúde multiatributos. Contudo, não requer que todos os estados de saúde sejam avaliados. Também permite que os domínios dentro de cada estado de saúde sejam separados e seu efeito analisado separadamente. Esta característica é importante porque possibilita conhecer como os avaliadores combinaram os diferentes domínios para chegar à avaliação global final.

1) Modelos explicitamente decompostos: o avaliador é solicitado a avaliar cada nível de determinado domínio enquanto todos os outros são mantidos constantes. Esta estratégia é derivada da clássica teoria da *Multiattribute Utility Theory* (MAU). Ela foi preconizada por Keeney & Raiffa (1993) que reconheceram a necessidade de expandir a teoria de utilidade de Von-Neumann e Morgenstern para as situações nas quais as decisões deveriam ser tomadas a partir de objetivos múltiplos e competitivos e não apenas de um objetivo único (KEENEY & RAIFFA, 1993). A teoria MAU está relacionada com a construção de funções de utilidade para múltiplos domínios. Especifica várias funções possíveis (aditiva, quase-aditiva e multilinear) e as condições de independência sobre as quais elas seriam apropriadas. A função de utilidade envolve três maiores passos (FROBERG & KANE, 1989a):

- a) Checar o pressuposto de independência
- b) Avaliar a função de utilidade em um único domínio
- c) Medir o índice de utilidade de alguns estados de saúde com múltiplos domínios a fim de determinar o valor total do estado de saúde para posteriormente decompor este valor nos valores de cada domínio. Por exemplo, se um estado contendo quatro domínios tem valor total 0,8, o valor de cada domínio poderia ser: domínio 1 = 0,2 + domínio 2 = 0,1 + domínio 3 = 0,3 + domínio 4 = 0,2.

O primeiro passo refere-se à independência entre os domínios (atributos) que estão incluídos no estado de saúde, ou seja, o efeito de um domínio (por exemplo,

saúde física) é independente do efeito dos outros domínios (por exemplo, saúde mental). Assim, o valor das preferências para cada nível do domínio “saúde física”, não depende do valor dos níveis do domínio “saúde mental” quando estes são mantidos constantes. Esta situação, onde não há interação entre os domínios, é chamado de modelo aditivo. Três condições devem ser satisfeitas para caracterizar um modelo como aditivo:

a) Independência dos índices de utilidade: o índice de utilidade de cada domínio é independente do índice de utilidade de outro domínio. Satisfazendo apenas esta condição, o modelo é denominado multilinear.

b) Independência mútua dos índices de utilidade: cada subgrupo de domínios tem seu índice de utilidade independente do seu complemento (o restante dos domínios). O modelo que satisfaça esta condição somada à condição descrita acima é chamado de modelo quasi-aditivo.

c) Independência do índice de utilidade aditiva: O índice de utilidade global do estado de saúde com múltiplos domínios é igual à soma dos índices de utilidade de cada domínio. A contribuição de cada domínio é independente dos valores dos outros domínios. Se esta condição é satisfeita juntamente com as outras duas, então o modelo é considerado aditivo.

Após determinação de qual modelo é apropriado, o investigador solicita ao avaliador que julgue sua preferência por cada nível de um domínio específico pressupondo que todos os outros domínios são mantidos constantes.

O terceiro e último passo é a medida das preferências por alguns estados de saúde multiatributos selecionados. Este procedimento é feito para a obtenção de constantes da escala. Estas constantes podem ser tidas como “pesos” para cada domínio.

Juntos, estes três passos representam o método de utilidade de multiatributos (MAU) que permite expressar índices de utilidade de estados de saúde com múltiplos domínios como uma função dos índices de cada domínio considerado separadamente.

2) Modelos decompostos por inferência estatística: tanto os modelos explicitamente decompostos como os modelos estatisticamente inferidos requerem um menor número de avaliações dos estados de saúde em relação aos modelos holísticos. Além disso, os modelos criados por inferência estatística, o chamado “método de medida funcional” tem a vantagem adicional de permitir um teste do

processo subjetivo subjacente através do qual os respondentes processam a informação, permitindo então também um processo de validação da medida (FROBERG & KANE, 1989a).

A medida funcional testa, simultaneamente, a teoria do processamento de informação e a validade dos valores da escala. De acordo com Anderson (1971), um construto subjetivo só pode ser medido no contexto de uma teoria válida. Este autor desenvolveu uma teoria explicando como o ser humano processa a informação e está exemplificada na figura 2:

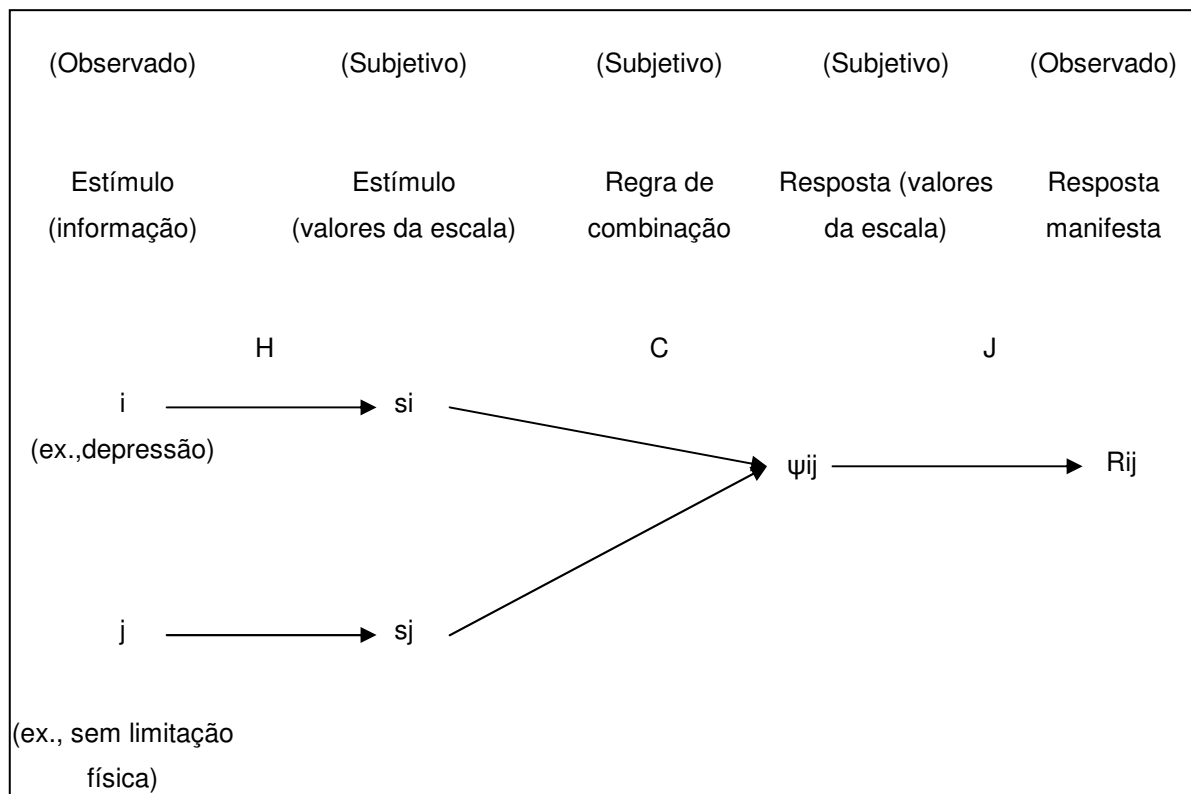


Fig. 2 - Teoria do processamento de informação (adaptado de FROBERG, 1989).

Aplicando esta teoria para o processo de avaliação de um estado de saúde com múltiplos domínios, poderíamos compreender como o indivíduo chega à resposta final através do seguinte mecanismo ilustrado na figura 2: apresentamos um estímulo ao respondente, que poderia ser, neste exemplo, um estado de saúde contendo dois domínios, um mental e um físico. Ele transforma cada “porção” desta informação (“depressão” e “sem limitação física”) em um “valor” subjetivo (si e sj) através de uma função (H). Após, o indivíduo utiliza uma regra de combinação (C) para transformar estes valores em uma resposta subjetiva (ψij). Finalmente,

transforma esta resposta subjetiva em uma resposta observada utilizando outra função (J). A medida então envolveria três processos simultâneos: 1) medir os valores do estímulo subjetivo em uma escala intervalar; 2) medir os valores da resposta subjetiva em uma escala intervalar; 3) encontrar a lei psicológica que relaciona os valores subjetivos do estímulo e da resposta. Na abordagem funcional, estas três questões seriam resolvidas conjuntamente. Esta tarefa requer o uso de um desenho fatorial. Tal técnica permite testar a lei que relaciona estes valores do estímulo e da resposta (regra de combinação C na figura 2). Se os dados se ajustam às predições do modelo, o estímulo subjetivo (s_i e s_j) e os valores da resposta (ψ_{ij}) poderiam ser derivados deste modelo. Suponhamos que no exemplo acima o domínio mental tenha cinco itens e o domínio físico seis itens. Em um desenho fatorial, todos os níveis do domínio mental podem ser combinados a todos os níveis do domínio físico para gerar 30 (5×6) possíveis estados de saúde. Os dados produzidos por um desenho fatorial são analisados usando procedimentos de análise de variância. Se os dados gerados pela avaliação de cada estado multiatributo (com múltiplos domínios) obedece às condições do modelo, este é aceito como uma descrição apropriada do processo de combinação, e o estímulo e o valor da resposta são separadamente derivados do modelo. O modelo aditivo é sustentado se nenhuma interação está presente. Se há interações, existem procedimentos disponíveis para determinar se estas interações podem ser descritas pelos modelos quase-aditivos ou multilineares descritos anteriormente. Quando não há interações, significa que os domínios são independentes, ou seja, o valor de um determinado domínio não depende dos valores dos outros domínios (FROBERG & KANE, 1989a).

Os problemas da medida funcional são basicamente logísticos. Em primeiro lugar, quando há muitos domínios e muitos níveis dentro de cada domínio, o número de avaliações de estados de saúde pode tornar a tarefa de obter um desenho fatorial completo quase impossível. Segundo, a aplicação destas técnicas exige experiência na área de desenho experimental e análise de variância, considerando que estas técnicas também são de difícil interpretação. O terceiro problema seria quando o número de domínios é grande, maior do que 6, pois isto dificulta a avaliação dos estados pelos indivíduos devido à grande quantidade de informação (FROBERG & KANE, 1989a).

Modelos de regressão múltipla

Outro método dentro dos modelos decompostos por inferência estatística é a regressão múltipla. Procedimentos de regressão têm sido utilizados para obter-se alguma compreensão da regra de combinação (C) da figura 2. Esta abordagem consiste em obter avaliações de estados de saúde multiatributos de uma amostra de indivíduos e após estimar os pesos subjetivos e os valores da escala através de um modelo de utilidade simples, geralmente o modelo aditivo, através de procedimento de regressão. Conclusões sobre a adequação do modelo são baseadas na magnitude do coeficiente de correlação múltipla. Se R^2 está entre 0,7 a 0,8, conclui-se que o grau de correspondência entre os índices de utilidade gerados pelo modelo e os valores obtidos das avaliações dos estados de saúde na amostra é alto. As análises de regressão assumem que os valores dos estímulos são conhecidos e que a resposta manifesta está em uma escala intervalar. O problema desta abordagem é que ela não testa a validade dos valores da escala. Enquanto a abordagem funcional incorpora a escala como uma parte da testagem da teoria subjacente ao processo de julgamento, as técnicas de regressão não oferecem esta possibilidade de validação. Os modelos de regressão simplesmente assumem que os valores são válidos e usam estes valores para testar a regra de combinação (FROBERG & KANE, 1989a).

Atualmente, vários instrumentos de amplo uso em diversos países utilizam modelos de regressão para estimar índices de utilidade, como o *Euro Quality of Life Instrument-5D* (EQ-5D, 1990) e *Short-Form 6D* (SF-6D), derivado do SF-36 (BRAZIER, ROBERTS, DEVERILL, 2002).

Para exemplificar a metodologia do uso de regressão múltipla, utilizaremos descrições do desenvolvimento do instrumento EQ-5D. O EuroQol é um instrumento cujo sistema descritivo é composto por cinco domínios (atributos), cada um com três níveis, como demonstrado na figura 3.

Domínios	Nível
Mobilidade	
Não tenho problemas em andar	1
Tenho alguns problemas em andar	2
Estou limitado a ficar na cama	3
Cuidados pessoais	
Não tenho problemas com meus cuidados pessoais	1
Tenho alguns problemas para me lavar ou me vestir	2
Sou incapaz de me lavar ou me vestir sozinho	3
Atividades usuais (trabalho, estudo, atividades domésticas, de família ou lazer)	
Não tenho problemas em desempenhar as minhas atividades habituais	1
Tenho alguns problemas em desempenhar minhas atividades habituais	2
Sou incapaz de desempenhar minhas atividades habituais	3
Dor/mal-estar	
Não tenho dores ou mal-estar	1
Tenho dores ou mal-estar moderados	2
Tenho dores ou mal-estar extremos	3
Ansiedade/depressão	
Não estou ansioso(a) ou deprimido(a)	1
Estou moderadamente ansioso(a) ou deprimido(a)	2
Estou extremamente ansioso(a) ou deprimido(a)	3

EuroQol-5D (EQ-5D, 2010).

Fig. 3 - Sistema descritivo do instrumento EuroQol-5D.

Para formar um estado de saúde, é selecionado um dos três itens de cada um dos cinco domínios, sendo possível gerar 243 estados de saúde ($3 \times 3 \times 3 \times 3 \times 3$). Um exemplo de estado de saúde gerado pelo EQ-5D está representado na figura 4.

Tenho alguns problemas em andar
 Não tenho problemas com meus cuidados pessoais
 Tenho alguns problemas em desempenhar minhas atividades habituais
 Tenho dores ou mal-estar moderados
 Estou moderadamente ansioso(a) ou deprimido(a)

Fig. 4 - Exemplo de estado de saúde gerado pelo EuroQol-5D.

Dos 243 estados, uma amostra é selecionada para ser avaliada por um grupo de indivíduos para que as preferências por cada estado desta amostra sejam medidas. Ao estimar relações paramétricas entre o sistema descritivo (figura 3) e estes valores de preferências medidos, é possível estimar os valores para os estados não avaliados (BUSSCHBACH et al., 1999). O modelo paramétrico deve prever os valores para todos os estados de saúde baseado nos escores do sistema descritivo. Em termos algébricos procura-se uma relação expressa por:

$$V = V(\vec{x})$$

onde V é o valor do estado de saúde, \vec{x} é o vetor da descrição do estado de saúde (isto é, os níveis dos estados de saúde demonstrados na figura 3) e $V(\vec{x})$ é uma função do valor.

Ao escolher o método de modelagem para estimar índices de utilidade de uma medida, algumas questões deverão ser consideradas (BUSSCHBACH et al., 1999):

1) Um modelo paramétrico requer valores em uma escala intervalar, pois não tendo esta propriedade não seria possível prever os valores dos outros estados que não foram medidos. Assim, assume-se que os valores gerados pelo instrumento utilizado para obter o valor V da fórmula acima (SG, *time trade-off* - TTO, ou escala análogo visual (EAV), serão abordados com mais detalhes na próxima seção) estão em escala intervalar. Por exemplo, o EQ-5D possui uma

versão que utiliza EAV, onde os entrevistados são orientados a alocar vários estados de saúde em posições na escala conforme suas preferências. Então, tomando-se como exemplo o estado de saúde apresentado acima e a regra do processamento de informação apresentada na figura 2, o indivíduo teria cinco “porções” de informação para processar, que seria um nível de cada um dos cinco domínios do instrumento. Ele atribui um valor subjetivo a cada um destes níveis (valores s_i e s_j na figura 2) e após combina a informação através da regra de combinação (C) para atribuir um valor ao estado de saúde como um todo (ψ_{ij}). Através da função J, o indivíduo manifesta sua resposta (R_{ij}), através da EAV (valor este que é o valor do ponto de alocação do estado na escala, por exemplo, 20). Porém, não é possível saber se a função que transforma o estímulo (estado de saúde) em resposta é determinada pelas propriedades da escala do estímulo (escores de cada nível no estado de saúde) ou pelas propriedades da escala da resposta (EAV). Uma vez que este problema de identificação é insolúvel, pressupõe-se uma relação linear entre o estímulo e a resposta a fim de que se possa gerar um modelo paramétrico para estimar todos os valores não medidos diretamente. Ao assumir este pressuposto, nada sabemos sobre as “reais” propriedades intervalares dos valores V . Então, não é possível testar a propriedade de escala da resposta apenas medindo os valores dos estados de saúde.

2) Utilização dos dados para construção do modelo: O modelo poderá ser gerado com os dados utilizados de duas formas: a) nível agregado: são utilizadas a média ou mediana dos valores para os estados de saúde avaliados por um grupo de indivíduos; b) nível individual: cada avaliação de cada estado de saúde é considerada como uma observação individual. Utilizando dados agregados, o número de graus de liberdade é determinado pelo número de estados de saúde avaliados empiricamente. Por outro lado, nas modelagens realizadas com dados em nível individual o grau de liberdade é determinado pelo número de sujeitos, que geralmente é maior do que o número de estados, o que aumentaria o poder estatístico para detectar parâmetros significativos.

Como os sujeitos não avaliam os mesmos estados de saúde, diferenças entre as preferências pelos estados de saúde podem ocorrer devido a diferenças entre os indivíduos e não por diferenças entre os estados. Avaliar este efeito das características do indivíduo só é possível na abordagem a nível individual e para isto é preconizado o uso de modelagens mais complexas, como modelos multi-nível ou

modelos de efeitos randômicos que atualmente tem sido introduzido nesta linha de pesquisa.

3) Modelagem: Seguindo tendo o EuroQol como exemplo, assume-se que cada estado de saúde pode ser caracterizado como um vetor $\vec{\chi}$ que representa os escores em cada uma das cinco dimensões e que existe uma relação linear entre este vetor e o valor dado ao estado de saúde na escala de medida. Esta relação linear pode ser descrita por uma função do valor $v(\vec{\chi})$ sobre a qual duas suposições são feitas: 1) Uma mudança mínima em qualquer uma das dimensões leva a uma mudança mínima no valor do estado de saúde; 2) Um escore melhor em cada dimensão deve corresponder a um valor maior do estado de saúde correspondente.

Há várias formas funcionais possíveis para esta “função do valor”. Uma das formas muito utilizadas na literatura é a forma aditiva, que implica que as contribuições das diferentes dimensões podem ser somadas para obter-se o valor do estado de saúde. Assume-se que tanto os escores de cada dimensão (1, 2 ou 3) como o valor do estado de saúde possuem propriedades intervalares. Então, pressupõe-se que o escore intermediário (2) é o ponto médio dos escores 1 e 3. Consideremos χ_i representando cada nível em cada dimensão.

$$V(X) = \alpha + \sum_{i=1}^5 \omega_i x_i + e$$

Equação retirada de Busschbach (1999).

Na equação acima, ω_i representa o “peso” atribuído a cada dimensão. Se a primeira dimensão (Mobilidade) tem um papel mais importante que a segunda (Cuidados pessoais), então o “peso” da primeira deve ser mais alto que o da segunda. O α é uma constante e e é um termo de erro. Como se assumiu que tanto v como χ_i estão em escala intervalar, a equação acima é simplesmente uma regressão múltipla onde v é a variável dependente, χ_i as variáveis explicatórias (independentes) e α o intercepto. Na regressão múltipla, o coeficiente de regressão β_i representa ω_i e pode ser usado como uma estimativa para estes pesos. Na modelagem, as variáveis independentes geralmente consistem de variáveis *dummy*

que representam os diferentes níveis de cada domínio dos instrumentos.

A última decisão a ser feita na geração de um modelo seria em relação à escolha do grupo de indivíduos para avaliação dos estados de saúde. Este tópico tem sido muito debatido na literatura e parece haver um consenso de que isto dependeria dos objetivos do estudo, a quem os valores interessam. Se os valores serão utilizados para tomada de decisão para alocação de recursos públicos em saúde, recomenda-se que a população geral seja a população avaliadora, o chamado ponto de vista social. Caso os pesquisadores estejam interessados em avaliar as preferências de pacientes, então este grupo deve ser selecionado (BUSSCHBACH et al., 1999).

Métodos para obter os valores para estados de saúde (escalas)

Acima foram descritos métodos para inferência de índices de utilidade após obtenção de valores empíricos para um grupo de estados de saúde. A seguir serão discutidas as técnicas para obter estes valores.

A medida de julgamentos subjetivos das pessoas é uma tarefa relacionada às ciências da psicométrica e psicofísica. Esta última está engajada com o estudo das maneiras pelas quais as pessoas percebem e fazem julgamentos sobre fenômenos físicos tais como a brilho de uma luz ou intensidade de um som. Desde o século 17, cientistas tem se empenhado em estabelecer relações matemáticas entre a intensidade dos estímulos e as sensações subjetivas, e descobriram que nem sempre esta relação é linear. Apesar disso, está claro que os humanos podem realizar estimativas numéricas consistentes dos estímulos sensoriais, até mesmo quando estes estímulos são abstratos. Então, os métodos da psicofísica foram adaptados para o estudo de julgamentos subjetivos quando não há nenhuma escala física, como calor, som, irradiação, incluindo as medidas de preferências humanas. Este é o campo da psicométrica (FROBERG & KANE, 1989b).

Não é escopo desta revisão discutir mais detalhadamente conceitos e técnicas psicométricas, mas citar como esta ciência foi adaptada para avaliação do desempenho das medidas de utilidade/preferências.

Para as medidas de utilidade, algumas distinções devem ser feitas em relação a outros tipos de medidas (FROBERG & KANE, 1989b):

a) a unidade de medida é um estado de saúde, ou seja, estamos medindo um

“estímulo” e não um indivíduo. É diferente de quando as pessoas respondem um instrumento que mede estado de saúde e um escore é atribuído a este indivíduo conforme sua resposta, localizando o indivíduo em algum ponto dentro da escala. Nas medidas de preferência, nós estamos interessados em localizar o estado de saúde dentro de um *continuum*.

b) o fato de que preferências por estados de saúde não são estímulos verificáveis. Não há nenhum padrão contra o qual comparar as respostas dos sujeitos, fazendo o processo de validação ocorrer muito mais através de um acúmulo de evidências do que por uma comparação definitiva.

c) tipo de escala: preferências por estados de saúde devem ser medidas em uma escala intervalar, pois esta é a maneira que permite o uso dos valores para análises de custo-efetividade. Uma escala intervalar permite intervalos iguais para diferenças entre estados de saúde. Por exemplo, se o índice de utilidade de um estado passou de 0,6 para 0,7 é equivalente a diferença do índice de outro estado quando mudou de 0,4 para 0,5, ou seja, 0,1.

d) medida direta X indireta: as técnicas de obter preferências por estados de saúde podem ser divididas em dois grandes grupos: técnicas diretas e indiretas.

Na abordagem direta, os valores são obtidos diretamente da escala de medida sem realizar nenhuma transformação. Por exemplo, se o valor da preferência por um estado de saúde for 0,8 medido através da técnica direta *standard gamble*, este será o número utilizado. Já nas medidas indiretas o indivíduo é instruído a realizar tarefas como, por exemplo, comparação de pares de estados de saúde e ele indica qual dos dois estados é o “preferido”. Isto gera respostas ordinais que deverão ser convertidas posteriormente para uma escala intervalar. Outra maneira indireta de medir preferências é através do preenchimento de um questionário contendo descrições de estados de saúde, e após as respostas dos indivíduos são convertidas em índices de utilidade através de um modelo estatístico conforme as técnicas descritas na seção anterior.

Métodos diretos

a) *Standard gamble*. A tradução para a língua portuguesa seria “jogo padrão”. Método clássico de medir preferências, sua base conceitual é derivada da teoria da decisão. Os primeiros autores a descrever esta técnica foram Von Neumann &

Morgestern e ela é baseada nos axiomas da teoria de utilidade, incorporando uma teoria da tomada de decisão sob condições de incerteza, como já descrito na seção 2.1.5.1. A essência da técnica é uma escolha apresentada ao indivíduo entre certo desfecho de ocorrência certa e uma “aposta” (*gamble*). A figura 5 ilustra os princípios da técnica:

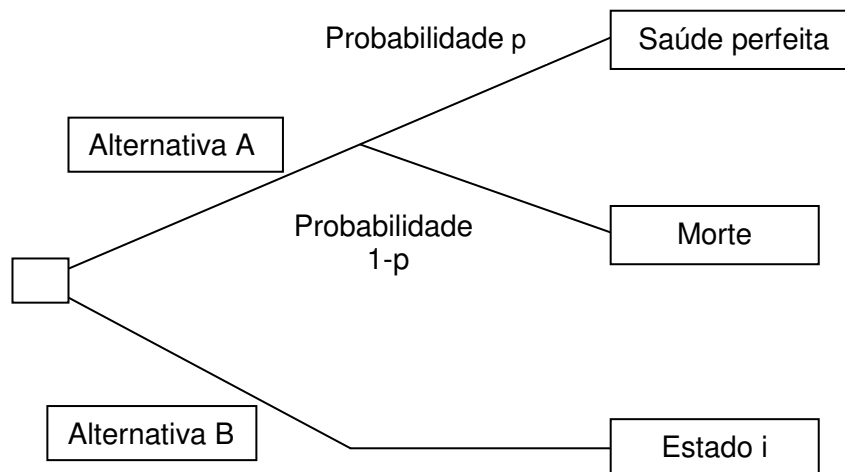


Fig. 5 - Representação gráfica do *standard gamble*.

Esta “aposta” consiste em perguntar ao indivíduo se ele aceitaria um tratamento (Alternativa A) que poderia ter dois possíveis desfechos: ficar com a saúde perfeita (probabilidade p) ou morrer imediatamente (probabilidade $1-p$). A alternativa B é o desfecho com ocorrência certa e geralmente consiste de um estado de saúde crônico (por exemplo, hemodiálise) pelo resto da vida. A probabilidade p é variada até que o indivíduo seja indiferente entre as duas alternativas (A e B). Neste ponto de indiferença o valor para o estado i é simplesmente o valor da probabilidade p .

Uma maneira intuitiva de compreender a técnica seria: se o estado i na figura 2 fosse um estado de saúde muito indesejável (por exemplo, tetraplegia), então o indivíduo seria indiferente em aceitar uma intervenção para mudar seu estado de saúde até mesmo com baixa probabilidade de voltar a um estado de saúde normal (por exemplo, 0,3). Então, poder-se-ia dizer que o valor da escala para paraplegia é baixo (0,3) (FROBERG & KANE, 1989b).

Variações nesta técnica são possíveis se o investigador está interessado em estados piores do que a morte ou estados temporários. A figura 6 ilustra como o SG

é aplicado para estados piores do que a morte.

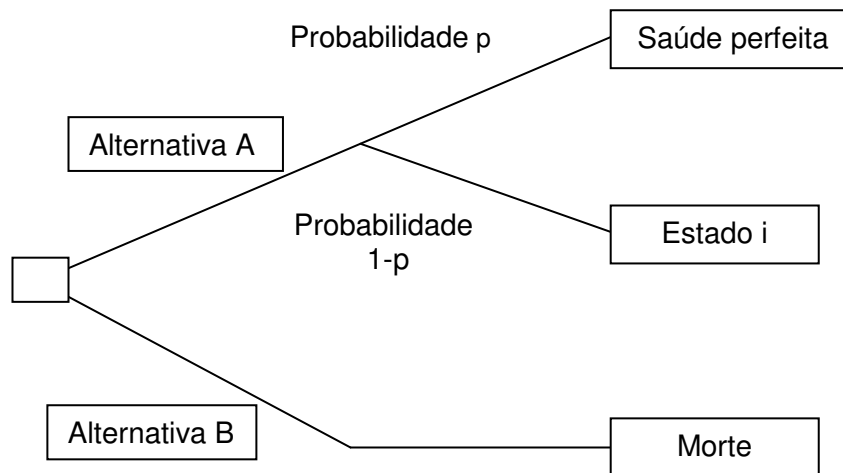


Fig. 6 - Representação gráfica do *standard gamble* para estados piores que a morte.

Nesta modalidade o desfecho com ocorrência certa seria morte imediata, enquanto a “aposta” seria saúde perfeita (probabilidade p) ou o estado i (probabilidade $1-p$). Uma maneira de apresentar as opções é solicitar que o indivíduo imagine que tem uma doença terminal que o levaria à morte caso não tratada. Se aceitar um tratamento, pode recuperar a saúde plena (probabilidade p) ou ficar em um estado crônico que seria considerado pior que a morte (probabilidade $1-p$). O valor para o estado i no ponto de indiferença seria dado por: $v_i = -p/(1-p)$. Intuitivamente, significa que se o estado i é um estado extremamente indesejável (por exemplo, estado vegetativo), um indivíduo provavelmente não escolheria a “aposta”, a não ser que a probabilidade de retornar a um estado de saúde normal fosse alta. Estados piores que a morte são representados por números negativos (FROBERG & KANE, 1989b).

Outra possibilidade é a aplicação para estados temporários como demonstrado na figura 7.

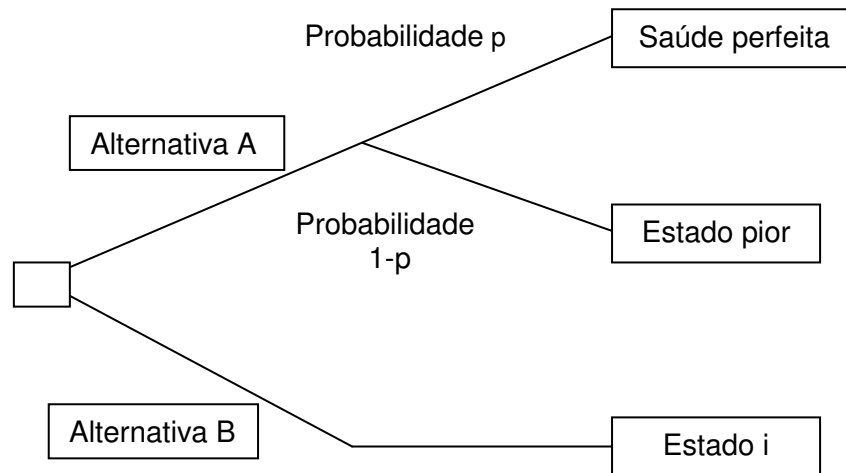


Fig. 7 - Representação gráfica do *standard gamble* para estados temporários.

Nesta alternativa o desfecho de ocorrência certa é o estado *i*, semelhante ao exemplo da figura 5, mas na “aposta” o desfecho morte imediata é substituído pelo pior estado de saúde. A fórmula para obter o valor do estado *i* seria: $v_i = p + (1-p) \cdot v_{\text{estado pior}}$.

Independente de qual das variantes é utilizada, esta técnica sempre envolve uma escolha entre um desfecho de certeza e um desfecho de incerteza (a “aposta”), onde o estado de saúde envolvido no desfecho de ocorrência certa é intermediário em termos de preferência em relação ao melhor e pior estado. Para tornar mais didática a tarefa de fazer com que os indivíduos façam escolhas pensando em probabilidades, o SG é realizado com a ajuda de um quadro onde as probabilidade *p* e *1-p* são demonstradas em cores diferentes e uma roda móvel vai variando estas probabilidades durante a tarefa. Esta variação é feita utilizando uma técnica denominada *ping-pong*, onde a probabilidade *p* e *1-p* inicia em 1 e 0, respectivamente, e vai variando de forma que, em cada momento, ora o melhor estado (*p*), ora o pior estado (*1-p*) são apresentados com mais chance de ocorrência, até o momento em que ambos tenham probabilidade de 0,5. Por exemplo: rodada 1. melhor estado 0,9/pior estado 0,1; rodada 2. melhor estado 0,1/pior estado 0,9; rodada 3. melhor estado 0,8/pior estado 0,2; rodada 4: melhor estado 0,2/pior estado 0,8 e assim sucessivamente, até que o respondente seja indiferente entre esta “aposta” e o desfecho de certeza (FURLONG et al., 1990).

Alguns autores consideram que “estudos de medidas de preferências de alta qualidade devem ser realizados com técnicas baseadas em escolhas como o *standard gamble*” (FURLONG et al., 1990). Esta afirmação deve-se ao fato de que os valores gerados pelo SG são considerados verdadeiros índices de utilidade por serem embasados na Teoria de Utilidade de Von Neumann & Morgenstern.

b) *Time trade-off* (TTO): Esta técnica foi desenvolvida por Torrance, Thomas e Sackett (1972), para uso específico na área da saúde e é considerada uma alternativa ao SG com administração mais simples. Neste método o sujeito também deve fazer uma escolha, mas neste caso a escolha é entre dois cenários de certeza, não havendo uma incerteza ou “aposta”. A técnica consiste em perguntar ao indivíduo quanto tempo de vida ele ou ela estaria disposto a desistir a fim de ficar em um estado de saúde perfeita comparado a um estado menos saudável. A figura 8 ilustra a técnica que consiste na seguinte sistemática: duas alternativas são oferecidas ao indivíduo: alternativa 1: o estado *i* pelo tempo *t*, seguida de morte. Alternativa 2: saúde perfeita por um tempo $x < t$ seguido de morte. O tempo *x* é variado até que o sujeito seja indiferente entre as duas alternativas. O valor do estado *i* neste ponto é dado por: $v_i = x/t$.

O TTO também pode ser realizado com alterações como citado para o SG, aplicando-o para estados piores que a morte e estados temporários (FROBERG & KANE, 1989b).

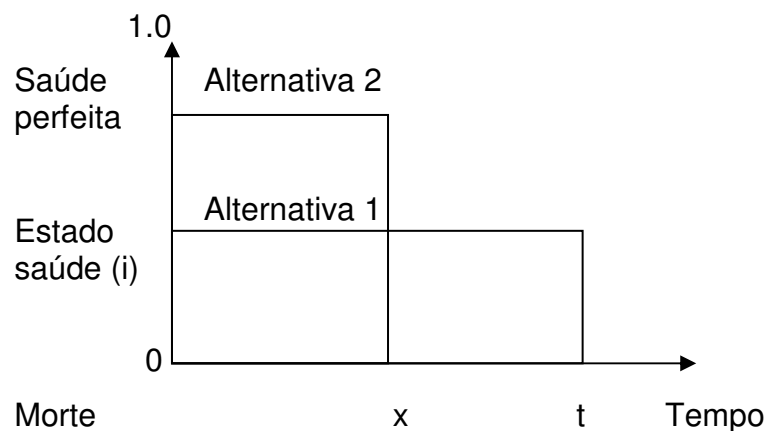


Fig. 8 - Representação gráfica do *Time trade-off*.

c) Escala análogo-visual: São escalas onde, em uma linha com valores extremos definidos (por exemplo, 0 e 100), o respondente deve identificar primeiramente qual o estado de saúde que para ele ou ela se localiza no valor mais baixo (neste caso 0) e no valor mais alto (100). Após, o sujeito pode avaliar uma série de estados intermediários e os aloca nesta linha entre os dois estados considerados “âncoras” conforme suas preferências, formando um ordenamento de estados. Geralmente, os pontos extremos são considerados saúde perfeita e morte, mas isto pode variar conforme a preferência do indivíduo. A fim de se obter uma escala intervalar, a forma de apresentação da tarefa deve ser feita a fim de que os indivíduos tenham pontos com igual distância para classificar os estados. Um exemplo é uma representação gráfica de um termômetro, onde os estados de saúde apresentados em cartões são colocados em cada um dos “graus” deste termômetro que está graduado de 0 a 100 em intervalos iguais (FROBERG & KANE, 1989b).

d) Outros métodos diretos de medir preferências são a magnitude de estimacão e o *willingness to pay* que serão apenas citados, pois atualmente são métodos menos utilizados na área da saúde.

Pesquisadores engajados com a teoria de decisão têm sido favoráveis ao uso do SG como método de eleição para medida de preferências pelo fato de ter sido construído sob as bases da teoria de utilidade e por implicar decisão sob condições de incerteza. Além disso, assume-se que o método gera valores em uma escala intervalar, embora esta suposição pareça ser mais conceitual do que empiricamente testada. Índices de utilidade derivados do SG podem apresentar o viés da aversão ao risco. Existe a possibilidade de que a resposta seja influenciada pela incapacidade do respondente de lidar com incerteza. Os economistas consideram a hipótese de que os indivíduos são avessos ao risco quando tem de escolher entre um desfecho de certeza favorável e uma “aposta” com probabilidade de outro desfecho favorável igual ou ainda maior. Por outro lado, pesquisas no campo da psicologia indicam que quando a escolha é entre um desfecho desfavorável e uma “aposta” com probabilidade maior de outro desfecho também desfavorável as pessoas aceitam mais correr riscos (FROBERG & KANE, 1989b). Considerando estas duas informações, Kahneman & Tversky (1982), estudaram prospectos de risco que envolvia os dois desfechos, positivo e negativo. O SG poderia servir como exemplo, pois a “aposta” é composta de saúde perfeita e morte. Estes autores

observaram que o prazer de um ganho é muito menos intenso que a dor de uma perda de tamanho similar. Estes achados sugerem que as pessoas geralmente preferem permanecer em um estado de saúde “menos que perfeito” do que arriscar ficar em um estado pior ou morrer. Assim, um estado de saúde deveria ser muito indesejável para que uma pessoa aceitasse um tratamento com risco moderado de morte. Esta atitude mais conservadora em relação ao risco pode fazer com que os índices de utilidade gerados sejam superestimados quando comparados aos índices gerados por outras técnicas que não envolvem risco.

Métodos indiretos

Os métodos diretos descritos acima para medida de preferências são instrumentos complexos e que consomem tempo, tornando sua aplicação prática difícil. Uma das alternativas que vem sendo amplamente utilizada atualmente é a aplicação dos instrumentos de QV baseados em preferências (*preference-based measures*) (DRUMMOND et al., 2005). Neste caso, os indivíduos respondem ao questionário e as respostas são transformadas em índices de utilidade através de fórmulas, ou algoritmos, construídos através de modelos de regressão conforme já discutido na seção 2.1.5.2. Os escores são baseados em preferências geralmente medidas na população geral através de um dos métodos diretos descritos acima (SG, TTO ou EAV) na fase de desenvolvimento e adaptação cultural dos instrumentos.

Os quatro instrumentos baseados em preferências mais amplamente utilizados internacionalmente são o *Quality of Well-Being* (QWB) (KAPLAN & ANDERSON, 1996), o *Health Utility Index* (HUI) (FURLONG et al., 2001) o EQ-5D (1990), do grupo EuroQol e o SF-6D (BRAZIER, ROBERTS, DEVERILL, 2002). O HUI, EQ-5D e o SF-6D já possuem versões em português brasileiro e alguns pesquisadores já aplicaram estes instrumentos no contexto nacional utilizando as fórmulas obtidas nos países de origem dos instrumentos para cálculos dos escores finais (SHIMODA et al., 2005; CAMPOLINA et al., 2009). Porém, tem sido reconhecido na literatura que as preferências de indivíduos por estados de saúde podem diferir de uma cultura para outra (BADIA et al., 2001; TSUCHIYA et al., 2002; GREINER et al., 2005; JOHNSON et al., 2005; LAM, BRAZIER, MCGHEE, 2008; BRAZIER et al., 2009), sendo recomendável que cada país tenha sua própria tabela

de valores de índices de utilidade para construção dos algoritmos. Este tipo de estudo ainda não foi realizado em nosso meio.

O QWB (KAPLAN & ANDERSON, 1996) classifica os indivíduos de acordo com quatro domínios: mobilidade, atividade física, atividade social, e um complexo de sintomas-problemas. A fórmula para cálculo dos escores é baseado em medidas do tipo EAV aplicadas em uma amostra da população geral da cidade de San Diego, EUA. É um questionário que exige longo tempo para administração, em média 14 minutos (DRUMMOND et al., 2005).

O EQ-5D (1990) foi desenvolvido por um consórcio de investigadores europeus, o grupo EuroQol. É composto de cinco domínios: mobilidade, autocuidado, atividades usuais, dor/desconforto e ansiedade/depressão, como exemplificado na seção 2.1.5.2. Cada domínio tem três níveis: nenhum problema, algum problema, problemas extremos. Gera 243 possíveis estados de saúde. As preferências para a construção do algoritmo para cálculo dos escores foram obtidas através da aplicação do TTO em uma amostra randômica de 3.000 indivíduos da população geral no Reino Unido. O algoritmo foi desenvolvido através de modelagem econométrica (DRUMMOND et al., 2005).

O SF-6D (BRAZIER, ROBERTS, DEVERILL, 2002) é um instrumento derivado do questionário de QV relacionada à saúde, o SF-36. Foi desenvolvido, em parte, porque o SF-36 é muito utilizado em um grande número de estudos e seria útil converter os escores do SF-36 para índices de utilidade, gerando QALYs para uso em análises econômicas. O SF-6D é composto por seis domínios: capacidade funcional, aspectos físicos e emocionais, aspectos sociais, dor, saúde mental e vitalidade. Cada domínio possui de quatro a seis níveis, podendo gerar 18.000 estados de saúde. O algoritmo para calcular os escores foi baseado em preferências medidas através do SG em uma amostra de 611 indivíduos da população geral da cidade de Sheffield, Inglaterra. O modelo também foi construído utilizando técnicas econométricas, como o EQ-5D. Este sistema permite que respostas ao SF-36 sejam convertidas para índices de utilidade.

O HUI (FURLONG et al., 2001) atualmente consiste de dois sistemas, HUI2 e HUI3. Cada um deles inclui um sistema descritivo e um algoritmo para calcular escores, que foi baseado em preferências medidas pelo SG em grupos populacionais no Canadá. É o único instrumento que foi baseado na Teoria de Utilidade Multiatributo (MAU) para a construção do modelo para cálculo dos escores finais.

2.1.5.3 Avaliação dos métodos baseados em preferências

Os instrumentos de avaliação de QV tradicionais são geralmente desenvolvidos e validados através das técnicas de psicometria clássica, ou outros métodos como a Teoria de Resposta ao Item (TRI), análises de Rasch e outras. Os instrumentos para medir índices de utilidade, por serem derivados da área da economia, nem sempre se encaixam nos critérios da psicometria para avaliação de “validade” e por este motivo acabam sendo muitas vezes negligenciados e considerados “inválidos” e irrelevantes (BRAZIER & DEVERILL, 1999). Como o construto “preferência” é difícil de ser avaliado objetivamente, há uma carência de estudos testando se os instrumentos disponíveis realmente estão medindo preferências das pessoas por estados de saúde. Os economistas assumem que o ser humano comporta-se conforme os axiomas da teoria de utilidade ao fazer escolhas e, portanto medidas baseadas nesta teoria, como o SG, são válidas por “natureza”.

Na ausência de métodos específicos para avaliação do desempenho das medidas de utilidade, pesquisadores têm adaptado as técnicas da psicometria.

Confiabilidade

Uma medida é considerada confiável se é relativamente livre de erro de medida. A confiabilidade determina o quanto a escala produz resultados consistentes. Em relação aos métodos para gerar preferências por estados de saúde, a confiabilidade pode ser avaliada por três métodos: a) confiabilidade intra-avaliador: consistência dos resultados gerados por um único avaliador quando o item é medido mais de uma vez; b) confiabilidade teste-reteste: estabilidade dos valores da escala entre dois períodos de tempo; c) confiabilidade inter-avaliador: consistência entre os valores de diferentes avaliadores em relação aos valores da escala.

Consistência interna, ou a consistência na resposta de item para item, não é aplicável aos métodos para gerar preferências. Não há razão para esperar correlações altas entre os estímulos, ou seja, os estados de saúde (FROBERG & KANE, 1989b).

A tabela 8 mostra dados disponíveis na literatura de coeficientes de confiabilidade dos métodos diretos de medida de preferências.

Tabela 8 - Confiabilidade dos métodos de medida direta de utilidade

Confiabilidade	SG	TTO	EAV
Intra-avaliador	0,77	0,77-0,88	0,70-0,94
Teste-reteste			
1 semana	0,80	0,87	0,77
4 semanas		0,81	
6 semanas		0,63-0,80	
1 ano	0,53	0,62	0,49
Inter-avaliador		0,75-0,77	

SG = *Standard gamble*;

TTO = *Time trade-off*;

EAV = Escala Análogo visual

Adaptado de Froberg (1989b)

Validade

Uma escala é válida se ela mede com precisão o que pretende medir. A validade pode ser de três tipos: de conteúdo, de critério e de constructo. A validade de constructo é um conceito mais amplo, e alguns autores consideram que ela inclui as outras duas. Aplicado às medidas de preferências por estados de saúde, a validade de conteúdo refere-se à adequação das descrições dos estados de saúde que representam o estímulo que será medido. A validade de conteúdo é alcançada por uma cuidadosa seleção dos domínios a serem incluídos no sistema descritivo e a apresentação dos estados com detalhamento suficiente (FROBERG & KANE, 1989b). Este critério é principalmente aplicado para os instrumentos baseados em preferências (método indireto), que possuem um sistema descritivo que gera os diferentes estados de saúde multiatributo, como o EuroQol-5D e o SF-6D. Para os métodos diretos de medida, este critério nem sempre se aplica, pois em muitos estudos o estado de saúde representa uma condição, por exemplo, hemodiálise, do que propriamente uma descrição de um estado de saúde.

Validade de critério não se aplica às medidas de preferências uma vez que

não existe um critério incorporado a preferências individuais, nem há tentativa de prever algum comportamento futuro com a medida de preferências. É uma variável abstrata, e não um fenômeno observável. Neste caso, a validação de constructo se aplica melhor. Muitas abordagens são possíveis para avaliar a validade de constructo, algumas têm sido aplicadas para as medidas de preferências, tais como examinar o grau pelo qual os resultados de diferentes métodos convergem e o grau pelo qual as relações previstas entre preferências e outras variáveis são empiricamente demonstradas. Estes métodos também representam uma alternativa para a falta de um padrão ouro em medidas de preferência contra o qual comparar as técnicas.

Alguns estudos compararam os métodos SG, *time-trade-off* e escala visual analógica. As correlações entre SG e *time-trade-off* foram de 0,56 a 0,65 conforme revisão realizada por Froberg & Kane (1989b). Estes dois métodos de medida parecem gerar valores diferentes de utilidade, mas há controvérsia na literatura em relação a qual deles gera valores mais altos. Alguns autores encontraram escores mais elevados utilizando o SG e este achado pode estar relacionado à aversão ao risco (FROBERG & KANE, 1989b). Dolan et al. (1996) encontraram um resultado oposto, os índices gerados pelo TTO foram mais elevados do que do SG. Doctor, Bleichrodt e Lin (2010) realizaram recentemente uma metanálise avaliando o efeito de vieses dos três métodos (SG, TTO e EAV) que podem influenciar as diferenças nos escores. Nenhum efeito significativo entre os escores do TTO e EAV foram observados. Em contraste, os escores da EAV foram significativamente mais baixos que o do SG, mas esta diferença foi anulada ao corrigir os escores do SG para vieses como aversão ao risco e efeito da forma de aplicação da técnica.

O SG tem sido considerado uma técnica válida por definição por estar embasada na teoria econômica de Von Neumann & Morgenstern. Contudo, é necessário um maior número de estudos, principalmente qualitativos, para testar se as preferências das pessoas por estados de saúde realmente comportam-se conforme os axiomas da teoria de decisão.

Responsividade

A responsividade de um instrumento de medida de QV é definida como a sua habilidade para detectar mudanças na saúde. Ela é geralmente estatisticamente

analisada utilizando medidas tais como o “tamanho de efeito” que é calculada dividindo-se a diferença nos escores pelo desvio padrão. De uma maneira geral, a medida com maior tamanho de efeito para uma dada mudança no estado de saúde, é considerada a melhor medida. Contudo, para as medidas utilizadas para avaliação econômica, as mudanças nas preferências do indivíduo são as que interessam. O tamanho de efeito de uma mudança clínica pode não refletir a importância desta mudança para os pacientes. Para as medidas de utilidade, responsividade deve representar a mudança nas preferências dos indivíduos pelo estado de saúde alcançado após uma intervenção (BRAZIER & DEVERILL, 1999).

Factibilidade

Para serem úteis, os métodos devem ser econômicos e aceitáveis para os respondentes. Tanto o SG como TTO são caros pela necessidade de material didático para a aplicação, treinamento extensivo de entrevistadores e um longo tempo de entrevista. Além disso, as pessoas podem considerar a tarefa difícil por ter de lidar com probabilidade e com risco no caso do SG. Este parece ter melhor desempenho com pessoas de mais alto nível educacional e quando a técnica de apresentação da tarefa é realizada com auxílio visual. Contudo, Luo et al. (2007) aplicaram o SG em diferentes grupos étnicos em Singapura e 75,9% da amostra consideraram o SG muito fácil ou fácil de compreender. Mais de 40% da amostra tinha menos de seis anos de estudo. Pesquisadores chineses incluíram no protocolo de aplicação do SG uma avaliação da experiência do entrevistado ao realizar a tarefa e apenas 21,4% da amostra sentiram-se irritados ou entediados com o exercício. Cerca de 36% consideraram a técnica fácil.

2.1.5.4 Efeitos da população estudada e do contexto de medida nos índices de utilidade

As pessoas geralmente possuem suas próprias visões e opiniões em relação a diversos aspectos da realidade, então é de se esperar que preferências por estados de saúde derivadas de indivíduos diferentes variem naturalmente entre si. Então, não é surpresa que os valores de utilidade medidos por determinado método

sejam diferentes em diversos subgrupos da população. No entanto, tem sido observado na literatura diferenças nos valores medidos no mesmo indivíduo. Considerando que todos os métodos deveriam medir “preferências”, seria esperado que mesmo técnicas diversas gerassem o mesmo valor quando medidos na mesma pessoa. Um dos fatores implicados nesta diferença seria a natureza do método, conforme discutido na seção anterior. Técnicas que envolvem risco podem gerar resultados diferentes das medidas que não exigem fazer escolhas em condições de incerteza.

Estas “inconsistências” vêm sendo estudadas por pesquisadores que buscam conhecer as variáveis relacionadas com valores inconsistentes em índices de utilidade.

Parece não haver um consenso do conceito de inconsistência na literatura, mas geralmente é considerada pelos autores de acordo com os seguintes critérios:

a) Os estados de saúde gerados pelos instrumentos baseados em preferências podem ser ordenados logicamente em relação a sua gravidade. Por exemplo, tomando o exemplo do EuroQol demonstrado na figura 3, um estado que tenha escore 1 ou 2 em todas as dimensões (12212) é melhor do que um estado que tenha um escore 2 ou 3 em todas elas (23323). Ou o estado 11111 (melhor) deve receber um valor maior do que o estado 333333 (pior). Quando esta regra é violada, considera-se que ocorreu uma inconsistência (BADIA, ROSET, HERDMAN, 1999; BRAVATA et al., 2005; LAMERS et al., 2006). A frequência com que esta situação ocorre parece variar de 17% a 49% (BRAVATA et al., 2005).

b) O indivíduo atribui o mesmo valor de utilidade a todos os estados quando estes são diferentes em termos de gravidade. Este tipo de inconsistência, denominada invariância na literatura, ainda é pouco estudado. Alguns autores relataram 24% de ocorrência de invariância (BRAVATA et al., 2005).

c) Badia, Roset e Herdman (1999) relataram a inconsistência de “critério”, quando métodos diferentes produzem um ordenamento diferente entre os estados avaliados. Estes autores compararam *ranking method*, *Time-trade off* e EAV.

Dolan & Kind (1996) sugeriram uma classificação didática de inconsistências em: a) primárias: aquelas relacionadas a fatores intrínsecos do indivíduo que está avaliando os estados de saúde; b) secundárias: relacionadas com a metodologia de medida.

Inconsistências primárias

Froberg & Kane (1989c) realizaram uma revisão dos aspectos que poderiam influenciar o julgamento de preferências, buscando evidências relatadas na literatura de quais variáveis podem causar diferenças em valores de índices de utilidade para um mesmo estado de saúde ou causar inconsistências. Os autores não encontraram nenhuma diferença sistemática em índices de utilidade que fosse associada com características demográficas, exceto idade e a experiência com uma doença. Contudo, outros autores relataram que, além da idade, o nível educacional também esteve associado com o nível de inconsistências (DOLAN & KIND, 1996; BADIA, ROSET, HERDMAN, 1999), ou seja, pessoas de maior idade e com menor nível educacional tem maior probabilidade de encontrar dificuldades em interpretar estados de saúde com múltiplos domínios e em compreender a tarefa de avaliação. Os autores salientam a possibilidade de ocorrer um viés sociodemográfico em estudos que excluem estes indivíduos, por incluir apenas sujeitos mais jovens e com nível educacional mais elevado. Esta é uma questão que suscita debate neste campo, uma vez que estes valores são muitas vezes utilizados para alocação de recursos públicos, levando a um paradoxo entre: a necessidade de valores consistentes para a construção de um modelo e ao mesmo tempo a necessidade de inclusão das preferências de todos os estratos populacionais.

Outros autores encontraram também associação entre raça/etnia e um maior tempo da entrevista com uma porcentagem maior de respostas inconsistentes (BRAVATA et al., 2005).

Inconsistências secundárias

Há vários aspectos do processo de medida de preferências que podem estar implicados no grau de inconsistência. Para fins didáticos, podemos dividir os fatores de acordo com a etapa realizada em um processo de avaliação de estados de saúde.

1. Forma de apresentação dos estados: A maneira pela qual os estados de saúde são apresentados e definidos podem ter influência no valor final dado ao estado (FROBERG & KANE, 1989c). Há evidências de que os valores dados a

estados de saúde quando descritos em forma de um cenário narrativo são menores do que aqueles valores para o mesmo estado quando descrito de uma forma mais padronizada (DOLAN & KIND, 1996). McNeil et al. (1982) demonstraram que as preferências dos entrevistados foram diferentes quando os cenários de decisão foram apresentados utilizando probabilidade de sobrevivência e probabilidade de morte. Inconsistências de valores para estados com múltiplos domínios poderiam também ser explicadas pela dificuldade em considerar estes estados de uma forma unidimensional (DOLAN & KIND, 1996). As pessoas podem atribuir “pesos” diferentes a cada domínio e julgar o valor do estado todo considerando uma ou outra dimensão. Para alguns indivíduos, limitações em funções mentais podem ser consideradas piores do que qualquer restrição no nível físico, por exemplo.

2. Técnica utilizada para obtenção dos valores: Diferentes métodos para medida de preferências parecem gerar graus diferentes de inconsistências. Estudos vêm demonstrando uma menor frequência de valores inconsistentes para os métodos simples de ordenamento como *ranking exercise* e escalas análogo-visuais quando comparados a *Time-Trade-off* (BADIA, ROSET, HERDMAN, 1999; LAMERS et al., 2006).

A experiência do entrevistador também foi demonstrada estar associada com inconsistências. Badia, Roset e Herdman (1999) encontrou duas vezes mais valores inconsistentes para entrevistas realizadas por entrevistador menos experiente.

2.2 APLICAÇÃO PRÁTICA DAS MEDIDAS DE UTILIDADE

Os índices de utilidade podem ser utilizados para uma simples descrição das preferências de grupos de pacientes com diferentes doenças e para avaliar uma possível mudança no valor destas preferências após uma intervenção. No entanto, a maior importância destas medidas reside na sua função como uma medida de efetividade para análises econômicas, as chamadas análises de custo-utilidade. Esta forma de avaliação está inserida no grupo das análises de custo-efetividade, mas foram desenvolvidas para lidar com algumas restrições destas últimas. Nas análises de custo-efetividade os desfechos são medidos em unidades específicas tais como milímetros de mercúrio de redução na pressão arterial, número de casos curados, anos de vida ganhos. O problema neste caso é que este desfecho mais

específico não permite comparações entre diferentes estratégias e não incorpora as preferências dos indivíduos por estes desfechos. Nas análises de custo-utilidade, o custo incremental de um programa é comparado ao ganho em saúde incremental atribuível ao programa, onde a melhora da saúde é medida pelos anos de vida ajustados para qualidade (QALY). Os resultados são expressos como custo por QALY ganho. Tendo uma unidade comum, vários programas diferentes podem ser comparados nas suas razões de custo-utilidade.

O QALY é uma medida que combina sobrevida e QV, esta representada pelos índices de utilidade.

Na prática, após obter um índice de utilidade utilizando as técnicas descritas nas seções anteriores, multiplica-se este índice pelos anos de vida gastos em determinado estado de saúde para obter-se o QALY. Como os índices de utilidade são medidos em escalas entre 0 e 1, onde 1 representa saúde perfeita, para simplificação os QALYs podem ser medidos em “anos em saúde perfeita” (um ano em perfeita saúde = 1 QALY, meio ano em perfeita saúde = 0,5 QALY). Podemos também calcular os QALYs ganhos em determinado tempo com uma intervenção. Por exemplo, consideremos que o índice de utilidade de diálise domiciliar é de 0,6. Para calcular quantos QALYs são ganhos por alcançar uma sobrevida de oito anos com diálise domiciliar, basta multiplicar $0,6 \times 8 = 4,8$ QALYs ganhos (DRUMMOND et al., 2005).

Uma alternativa ao uso do QALY é o *Disability-Adjusted Life-Years* (DALY), uma medida desenvolvida pela OMS inicialmente para o estudo *Global Burden of Disease and Injury* e atualmente tem sido recomendada por esta instituição para análises de custo-efetividade (DRUMMOND et al., 2005). O DALY é conceitualmente similar ao QALY, mas algumas diferenças são: 1) a expectativa de vida utilizada para o QALY depende do contexto do estudo. A utilizada para o DALY é constante e tem como referência a maior expectativa de vida descrita, que é a de mulheres japonesas; 2) os valores utilizados no cálculo de QALYs são baseados em preferências, tanto do público geral como de populações específicas. Os valores para calcular DALYs são escores obtidos por técnicas de Permuta pessoal (*person trade-off*) gerados através de um consenso de profissionais da área da saúde ocorrido em Genebra em 1995; 3) Ambas as escalas, de QALY e DALYs são representadas por dois pontos extremos, 0 = morte e 1 = saúde perfeita. Porém, os valores utilizados no DALY provêm de um sistema onde existe apenas sete estados

de saúde além de morte e saúde perfeita; 3) Diferente do QALY, o DALY utiliza um ajuste para a idade, com um “desconto” no valor para os extremos etários (mais jovens e mais velhos) (DRUMMOND et al., 2005).

A abordagem do DALY ainda tem sido alvo de intenso debate, por isto ainda não é uma medida mais amplamente utilizada em análises econômicas. Ela vem sendo mais recomendada pela OMS para uso nas análises de custo-efetividade generalizadas, que é uma abordagem desenvolvida pela própria OMS para avaliar uma ampla gama de intervenções para doenças crônicas no contexto de um orçamento fixo (DRUMMOND et al., 2005).

As recomendações do Painel de Custo-Efetividade em Saúde e Medicina, publicadas em 1996 (GOLD et al., 1996), para aplicação das medidas de QV em análises econômicas ainda estão sendo utilizadas atualmente como uma diretriz para uso deste tipo de instrumento. Este documento pode ser uma referência para pesquisadores interessados em desenvolver e validar medidas de utilidade e para aqueles envolvidos na execução de análises de custo-utilidade. As recomendações estão descritas abaixo:

1. Para as análises do caso-referência (*reference case*), a incorporação dos desfechos de morbidade e mortalidade deve ser combinada em uma única medida utilizando QALYs (caso-referência, para estudos de custo-efetividade, é considerado um padrão de práticas metodológicas que um analista deve procurar seguir em um estudo de custo-efetividade. Representa a análise de referência para comparações através dos estudos).

2. Em geral, como os anos de vida salvos ou estendidos por uma intervenção não serão vividos em saúde perfeita, um ano de vida salvo não será 1 QALY. Um ano de vida salvo deverá ser ajustado para a QV (anos de vida x índice de utilidade da condição de saúde em questão).

3. Para satisfazer o conceito de QALY, os índices de ajuste para QV devem ser: baseados em preferências, com características intervalares e medidos ou transformados em uma escala onde os pontos de referências são morte = 0 e saúde perfeita = 1.

4. As preferências por estados de saúde mais apropriadas para uso no caso-referência devem ser aquelas medidas na comunidade.

5. Quando as preferências da comunidade são utilizadas e o programa sendo analisado é relacionado a uma doença ou condição específica, é importante

realizar análises de sensibilidade com dados contendo informação das preferências dos pacientes portadores desta condição.

6. Se valores de preferências entre diferentes subgrupos divergem e esta diferença pode afetar a razão de custo-efetividade, a análise do caso-referência deve prover esta informação e deve ser conduzida análise de sensibilidade separadamente.

7. A QV pode ser afetada por variáveis como idade, sexo, raça ou situação socioeconômica das pessoas. É recomendado incluir estas variações nas análises de sensibilidade a fim de indicar como elas afetam o resultado final.

8. A análise de custo-efetividade deve ser baseada em um instrumento de QV relacionada à saúde que inclua os domínios importantes para a população-alvo do estudo. Se a análise é planejada para uso como caso-referência, a medida de preferência escolhida deve ser genérica a fim de possibilitar comparações entre programas.

9. Nas análises de caso-referência, a QV relacionada à saúde deve ser medida por um instrumento que incorpore os efeitos da morbidade na produtividade e no lazer dos indivíduos.

10. Consequências financeiras relacionadas a mudanças no estado de saúde, incluindo o valor do tempo de doença para os pacientes, devem ser representadas no denominador de uma razão de custo-efetividade através de valores das preferências destes pacientes. Efeitos financeiros e temporais que não estão relacionados ao estado de saúde devem ser incluídos no numerador.

3 OBJETIVOS

3 OBJETIVOS

3.1 OBJETIVOS PRINCIPAIS

1. Adaptação cultural, em uma amostra da população da cidade de Porto Alegre, da medida de QV baseada em preferências SF-6D, originalmente desenvolvida no Reino Unido

2. Aplicação em amostra da população geral dos instrumentos genéricos de QV WHOQOL-breve e SF-36 a fim de disponibilizar escores normativos destes instrumentos.

3.2 OBJETIVOS SECUNDÁRIOS

1. Descrever a metodologia para medidas de preferências (índices de utilidade), uma vez que esta é uma técnica ainda pouco conhecida no cenário nacional

2. Avaliar o desempenho da aplicação de técnicas complexas como as utilizadas para medir preferências na população de estudo e as associações dos resultados com características sociodemográficas

3. Estimar preferências (utilidade) para diferentes estados de saúde em amostra da população geral através da aplicação do questionário SG.

4 REFERÊNCIAS BIBLIOGRÁFICAS

4 REFERÊNCIAS BIBLIOGRÁFICAS

Anderson NH. Integration theory and attitude change. *Psychological Review* 1971; 78(3): 171-206.

Badia X, Roset M, Herdman M. Inconsistent responses in three preference-elicitation methods for health states. *Soc Sci Med* 1999; 49(7): 943-950.

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.

Bowling A. *Measuring health*. Open University Press, Filadélfia 1997.

Bravata DM, Nelson LM, Garber AM, Goldstein MK. Invariance and inconsistency in utility ratings. *Med Decis Making* 2005;25(2):158-67.

Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ* 1999;8(1):41-51.

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.

Brazier JE, Fukuhara S, Roberts J, Kharroubi S, Yamamoto Y, Ikeda S, Doherty J, Kurokawa K. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol* 2009;62(12):1323-31.

Busschbach JJ, McDonnell J, Essink-Bot ML, van Hout BA. Estimating parametric relationships between health description and health valuation with an application to the EuroQol EQ-5D. *J Health Econ* 1999;18(5):551-571.

Campbell A, Converse PE, Rodgers WL. *The quality of american life: perceptions, evaluations and satisfactions*. Russel Sage Foundation, Nova York 1976.

Campolina AG, Bortoluzzo AB, Ferraz MB, Ciconelli RM. Validity of the SF-6D index in Brazilian patients with rheumatoid arthritis. *Clinical and Experimental Rheumatology* 2009;27:64-71.

Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras de Reumatol* 1999;39(3):143-50.

Cohen J. *Statistical Power analysis for the behavioral sciences*. Lawrence Erlbaum, Hillsdale 1988.

Doctor JN, Bleichrodt H, Lin HJ. Health utility bias: a systematic review and meta-analytic evaluation. *Med Decis Making* 2010;30(1):58-67.

Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996;15(2):209-31.

Dolan P, Kind P. Inconsistency and health state valuations. *Soc Sci Med* 1996;42(4):609-15.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. *Methods for the economic evaluation of health care programmes*. Oxford University Press, Oxford 2005.

EuroQol. A new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy* 1990;16(3):199-208.

EuroQol. EQ-5D. A standardized instrument for use as a measure of health outcome. EuroQol Group [serial online] 2010a. Available from: <http://www.euroqol.org/eq-5d/eq-5d-versions/alternative-modes.html>. Access: 06 mar. 2010.

Fayers PM, Machin D. *Quality of Life. The assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons, West Sussex, England 2007.

Feeny DH, Torrance GW. Incorporating utility-based quality-of-life assessment measures in clinical trials. Two examples. *Med Care* 1989;27(3 Suppl):S190-S204.

Fleck MPA. Avaliação de qualidade de vida. In: *Depressões em Medicina Interna e em outras condições médicas*. (Eds. Fráguas RJ, Figueiró JAB). Atheneu: São Paulo, 2000;Primeira:33-43.

Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, Pinzon V. Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref. *Rev Saude Publica* 2000;34(2):178-83.

Fleck MPA. Problemas conceituais em qualidade de vida. In: A avaliação de qualidade de vida. Guia para profissionais da saúde. (Ed. Fleck MPA). Artmed: Porto Alegre, 2008;Primeira:19-28.

Froberg DG, Kane RL. Methodology for measuring health-state preferences-I: Measurement strategies. *J Clin Epidemiol* 1989a;42(4):345-54.

Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: Scaling methods. *J Clin Epidemiol* 1989b;42(5):459-71.

Froberg DG, Kane RL. Methodology for measuring health-state preferences-III: Population and context effects. *J Clin Epidemiol* 1989c;42(6):585-92.

Furlong W, Feeny D, Torrance G, Barr R, Horsman J. Paper #90-9 Guide to design and development of Health-State Utility Instrumentation. 1990. Hamilton, Ontario, Canada, McMaster University. CHEPA Working Paper Series. Ref Type: Serial (Book,Monograph)

Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001;33(5):375-84.

Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994;272(8):619-26.

Gold MR, Patrick DL, Torrance GW, Fryback DG, Hadorn DC, Kamlet MS, Daniels N, Weinstein MC. Identifying and valuing outcomes. In: Cost-effectiveness in health and Medicine. (Eds. Gold MR, Russell L, Siegel JE, Weinstein MC). Oxford University Press, New York, 1996;82-134.

Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 2005;6(2):124-30.

Gurin G, Veroff J, Feld S. Americans view their mental health. Basic Books, Nova York 1960.

Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118(8):622-29.

Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. *JAMA* 1997;277(15):1232-7.

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.

Kahneman D, Tversky A. The psychology of preference. *Sci Am* 1982;246:160-73.

Kaplan R, Anderson JP. The general health policy model: an integrated approach. In: *Quality of life and pharmacoeconomics in clinical trials*. (Ed. Spilker B). Lippincott-Raven, Philadelphia, 1996; Second:309-22.

Keeney RL, Raiffa H. *Decisions with multiple objectives. Preferences and Value tradeoffs*. Cambridge University Press, Cambridge 1993.

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.

Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making* 2006;26(2):173-81.

Lepelge A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997;278(1):47-50.

Luo N, Wang Q, Feeny D, Chen G, Li SC, Thumboo J. Measuring health preferences for Health Utilities Index Mark 3 health states: a study of feasibility and preference differences among ethnic groups in Singapore. *Med Decis Making* 2007;27(1):61-70.

McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32(1):40-66.

McNeil BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 1982;306:1259-62.

Measuring effectiveness and cost-effectiveness: the QALY. National Institute for Health and Clinical Excellence (NICE) [serial online] 2010. Available from: <http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenessstheqaly.jsp>. Access: 12 fev. 2010

Mosteller F, Falotico-Taylor. *Quality of life and technology assessment*. monograph of the council on health care technology. Mosteller F and Falotico-Taylor. 1989. Washington, National Academy Press.

Noerholm V, Groenvold M, Watt T, Bjorner J B, Rasmussen N A, Bech P. Quality of life in the Danish general population-normative data and validity of WHOQOL-BREF using Rasch and item response theory models. *Qual Life Res* 2004;13(2):531-40.

Ohaeri JU, Awadalla AW, Gado OM. Subjective quality of life in a nationwide sample of Kuwaiti subjects using the short version of the WHO quality of life instrument. *Soc Psychiatry Psychiatr Epidemiol* 2009;44(8):693-701.

Patrick DL, Bush JW, Chen MM. Toward an operational definition of health. *J Health Soc Behav* 1973;14(1):6-23.

Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27(3 Suppl):S217-S232.

Patrick DL, Erickson P. Applications of health status assessment to health policy. In: *Quality of life and pharmacoeconomics in Clinical Trials*. (Ed. Spilker B). Lippincott-Raven, Philadelphia, 1996; Second:717-27.

Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B. Patient-reported outcomes: conceptual issues. *Value Health* 2007;10 Suppl 2:S66-S75.

Scoggins JF, Patrick DL. The use of patient-reported outcomes instruments in registered clinical trials: evidence from Clinical Trials. *Gov. Contemp Clin Trials* 2009; 30(4):289-92.

Shimoda SCB, Horsman J, Furlong W, Lopes LF, Seber A, Barr RD. Translation and cultural adaptation of Health Utilities Index (HUI) Mark 2 (HUI2) and Mark 3 (HUI3) with application to survivors of childhood cancer in Brazil. *Qual Life Res* 2005;14(5): 1407-12.

Swenson JR. Quality of life in patients with coronary artery disease and the impact of depression. *Curr Psychiatry Rep* 2004;6(6):438-45.

The World Health Organization Quality of Life Assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med* 1995;41(10):1403-09.

The World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. *Psychol Med* 1998a; 28(3):551-8.

The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 1998b;46(12):1569-85.

Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7(2):118-33.

Torrance GW. Toward a utility theory foundation for health status index models. *Health Serv Res* 1976;11(4):349-69.

Tsevat J. What do utilities measure? *Med Care* 2000; 38(9 Suppl):II160-II164.

Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisashige A, Tamura M. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002;11(4):341-53.

Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton 1944.

Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.

Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. *International Quality of Life Assessment. Int J Technol Assess Health Care* 1995;11(3):525-51.

Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976)* 2000;25(24) 3130-9.

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996;276(15):1253-8.

Wenger NK. Improvement of quality of life in the framework of cardiac rehabilitation. In: *Clinical Cardiac Rehabilitation: A Cardiologist's Guide*. (Eds. Pashkow FJ, Dafoe WA). Williams & Wilkins, Baltimore, 1999; Second:43-51.

Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273(1):59-65.

ARTIGO 1

**ESTIMATIVA DE ÍNDICES DE UTILIDADE A PARTIR
DO SF-36 NO BRASIL:
DESCRIÇÃO DA METODOLOGIA**

ESTIMATIVA DE ÍNDICES DE UTILIDADE A PARTIR DO SF-36 NO BRASIL: DESCRIÇÃO DA METODOLOGIA

Luciane N. Cruz, MD, MSc ¹ , Marcelo P. Fleck, MD, ScD² , Suzi A. Comey, ScD³,
Michele R. Oliveira⁴, Carisi A. Polanczyk, MD, ScD ^{1, 5}

Artigo em submissão para a Revista de Saúde Pública

Programa de Pós-Graduação em Epidemiologia¹, Programa de Pós-Graduação em Psiquiatria², Departamento de Estatística³, Universidade Federal do Rio Grande do Sul; Graduação em Biologia, Universidade do Vale dos Sinos⁴; Serviço de Cardiologia, Hospital de Clínicas de Porto Alegre⁵

RESUMO

Recomendações internacionalmente estabelecidas para estudos de análise econômica em saúde preconizam a utilização de medidas de utilidade para a análise de custo-efetividade. Um dos métodos para obtenção destes valores são as medidas de qualidade de vida chamadas “medidas baseadas em preferências”. Recentemente instrumentos genéricos para medir qualidade de vida relacionada à saúde como o MOS Short-Form 36 (SF-36), por exemplo, tem sido transformados em medidas de utilidade, gerando uma medida denominada SF-6D. Entretanto, o emprego de tais instrumentos para outras populações requer uma série de etapas para sua validação. Neste artigo descrevemos a metodologia empregada por pesquisadores brasileiros para medir utilidade em uma amostra da população geral utilizando estes instrumentos. A partir dos valores obtidos, foi construído um algoritmo brasileiro para o SF-6D. A técnica para medidas de preferências ainda é pouco difundida nacionalmente e o presente artigo tem o objetivo de descrever a metodologia utilizada no projeto.

ABSTRACT

International recommendations for economic evaluation studies in health advocate the use of utility index for cost-effectiveness analysis. One of the several methods to obtain utilities is the use of preference-based measures, a kind of generic quality of life instrument. Recently a generic instrument to measure health related quality of life, MOS Short-Form 36 (SF-36) has been transformed into preference-based instrument, the SF-6D. However, the cross-cultural adaptation of such measures demands a multistage protocol. In this article, the methodology applied to measure utilities in a general Brazilian population sample is described. From the values obtained, a Brazilian algorithm for SF-6D was built. The technique for measuring preferences is still little known nationally and this article aims to describe the methodology used in the project.

INTRODUÇÃO

A avaliação de qualidade de vida vem se tornando nas últimas décadas uma importante medida de desfecho utilizada em ensaios clínicos e em estudos epidemiológicos, medindo o impacto de diferentes doenças e seu respectivo tratamento na vida dos indivíduos e também como medida de efetividade em análises econômicas.

Os instrumentos para avaliar qualidade de vida podem ser divididos em dois grupos: a) medidas descritivas, ou psicométricas; b) métodos baseados em preferências, que geram *utilidade*, um único índice que representa a preferência dos indivíduos por determinado estado de saúde.¹⁴ Estes índices são utilizados para gerar os QALYs (*Quality Adjusted Life Years*), que são as medidas de desfecho recomendadas para uso em estudos de custo-efetividade.²³ Entretanto, os métodos usualmente aceitos para a mensuração de preferências, os chamados métodos diretos de medida, como o *standard gamble* (SG) (Jogo-padrão) e *time trade-off* (TTO)(Permuta temporal)¹², são de difícil aplicação em estudos clínicos. Atualmente, instrumentos que agregam estes dois conceitos têm sido desenvolvidos e validados, a fim de que as medidas psicométricas possam ser também utilizadas em análises econômicas.¹⁶ Estes têm sido denominados de métodos indiretos ou medidas baseadas em preferências. O desenvolvimento destas medidas compõe-se de 3 etapas: 1) composição de um sistema descritivo que irá gerar diferentes estados de saúde em vários domínios; 2) seleção de uma amostra de estados de saúde para quantificação das preferências de um grupo de indivíduos através de uma das

técnicas citadas acima, SG, TTO ou escala análogo visual; 3) construção de um algoritmo através de modelos de regressão utilizando os valores obtidos na etapa 2. Esta equação tem como objetivo estimar os valores de preferências para todos os outros estados de saúde gerados pelo instrumento, sem a necessidade de que todos sejam avaliados.

Os instrumentos baseados em preferências mais conhecidos internacionalmente são o EuroQol-5D(18), o Health Utility Index¹⁰ e o SF-6D.⁴ Este último foi desenvolvido no Reino Unido e constitui-se por um sistema capaz de gerar índices de utilidade a partir do instrumento psicométrico SF-36.²² Uma primeira versão do instrumento foi publicada em 1998³, sendo revisada em 2002.⁴

O sistema descritivo SF-6D em português baseado na primeira versão do protocolo britânico já é encontrado na literatura⁶, mas ainda não há disponível no Brasil um algoritmo construído com base em índices de utilidade medidos na nossa população. Há evidências na literatura de que existem diferenças nas preferências entre culturas diferentes^{5;11;15;20}, sendo desejável que cada país possa obter seu próprio sistema de valores. Considerando as recomendações internacionalmente estabelecidas em relação à necessidade de utilização de parâmetros de utilidade para a análise de custo-efetividade^{9;17;23}, a carência destes índices determina um grande hiato para a pesquisa em saúde no país.

Em colaboração com Brazier et al.⁴ foi desenvolvido um estudo para replicar o protocolo britânico buscando disponibilizar um algoritmo para cálculo de índices de utilidade a partir do SF-36 baseado em preferências medidas em uma amostra da população da cidade de Porto Alegre.

O objetivo do presente artigo é descrever a metodologia empregada para a medida de índices de utilidade, uma vez que os instrumentos e técnicas utilizados

são ainda pouco conhecidos no contexto nacional, justificando sua apresentação à comunidade científica brasileira. Não é do escopo deste manuscrito descrever com maiores detalhes o processo de seleção da amostra por não ser diferente de técnicas usuais, tampouco a metodologia de análises dos dados, que estarão descritas em um artigo específico.

DESCRIÇÃO DA METODOLOGIA

Este projeto foi realizado em parceria com pesquisadores da Universidade de Sheffield, Inglaterra, com intercâmbio de pesquisadores da Universidade Federal do Rio Grande do Sul visando o planejamento e treinamento para aplicação dos instrumentos. O protocolo foi o mesmo utilizado no Reino Unido e também já aplicado em Portugal¹¹, China¹⁵ e Japão⁵.

Alguns conceitos serão brevemente revisados para melhor compreensão da terminologia utilizada no decorrer do artigo.

A teoria de *utilidade* foi desenvolvida no campo da economia como uma maneira de quantificar e analisar a maneira pela qual as pessoas fazem escolhas. Em 1944, Von Neumann e Morgenstern²¹, publicaram sua teoria de tomada de decisão racional sob condições de incerteza, a chamada “Teoria de Utilidade de Von Neumann e Morgenstern”.

Estes autores desenvolveram um modelo normativo em relação ao comportamento das pessoas ao tomar decisões, ou seja, como indivíduos racionais deveriam fazer escolhas quando confrontados com situações de incerteza. O uso desta teoria foi disseminado naturalmente para a área da saúde, gerando os chamados “índices de utilidade para estados de saúde”.¹⁹

O termo *utilidade* para as ciências da saúde, portanto, refere-se à preferência dos indivíduos por um desfecho, que neste caso é um estado de saúde. As medidas de utilidade avaliam o valor ou conveniência de um determinado estado de saúde.

Os termos “utilidade”, “valor” e “preferência” são muitas vezes usados como sinônimos, mas existem diferenças entre eles. Preferência é um termo “guarda-chuva” que descreve um conceito mais geral e utilidade e valor são tipos diferentes de preferências, dependendo de como a medida é realizada. Utilidade seria um termo específico para descrever preferências medidas com instrumentos baseados na Teoria de Von Neumann e Morgenstern, que implicam em uma escolha feita sob condições de incerteza.²¹ Este instrumento é o *standard gamble* e sua técnica será descrita mais adiante. No entanto, apesar de haver esta diferença teórica entre os termos, é comum na literatura os autores atribuírem o termo utilidade para os valores obtidos através de outras técnicas.

Independente da base teórica das preferências, elas devem ser mensuradas através de uma escala com as seguintes características: a) ser intervalar; b) ter intervalos iguais de forma que uma mudança de, por exemplo, 0,2 para 0,3 tenha o mesmo significado de uma mudança de 0,8 para 0,9; c) serem medidas em uma escala entre 0 e 1 onde 0 representa morte e 1 um estado de saúde perfeita.⁸

Derivação do SF-6D a partir do SF-36

O SF-36 é um instrumento que mede estado de saúde em 8 domínios: capacidade funcional, aspectos físicos, dor, estado geral de saúde, vitalidade, aspectos sociais, aspectos emocionais e saúde mental (Quadro 1). Cada domínio é formado por itens que questionam o indivíduo sobre sua percepção a respeito de sua própria saúde.

Contudo, instrumentos de qualidade de vida relacionada à saúde com a estrutura do SF-36, do ponto de vista dos economistas e da teoria de Von Neumann e Morgenstern²¹, não são adequados para serem incorporadas como uma medida de desfecho em estudos de custo-efetividade pelos seguintes aspectos: 1) os instrumentos geralmente não produzem um único escore de qualidade de vida, mas vários escores em diferentes domínios. Não é possível comparar uma melhora em um domínio com outro. Para as análises econômicas é necessário um único número que possibilite comparações entre diferentes intervenções que levam a desfechos diversos; 2) os escores dos instrumentos não são baseados nas preferências dos indivíduos pelos vários desfechos possíveis; 3) os escores dos instrumentos psicométricos não são calibrados em uma escala entre 0 (morte) e 1 (saúde perfeita), impossibilitando a combinação com sobrevida para gerar o QALY (*Quality Adjusted Life Years*).⁸

Portanto, o SF-36 foi reduzido em tamanho e complexidade para que fosse possível aplicar métodos para gerar preferências.

A derivação de itens a partir do SF-36 para gerar o SF-6D seguiu rigorosamente a mesma metodologia do protocolo britânico⁴, utilizando a versão em português do SF-36⁷. Os autores britânicos basearam o processo de seleção de itens nas análises fatoriais realizadas pelos autores do SF-36 original²², estabelecendo a importância de cada item para os escores globais de cada domínio.

O número de domínios do SF-36 foi reduzido de 8 para 6, obedecendo-se os seguintes passos: 1) O domínio “Estado geral de saúde” foi suprimido, pois não faria sentido manter um escore de saúde geral quando o objetivo do novo instrumento é gerar um escore global; 2) O domínio “Aspectos físicos” foi combinado com o de “Aspectos emocionais” para formar um único domínio, sem perda de distinção entre

estes dois aspectos; 3) Os seguintes itens foram selecionados do SF-36, em cada domínio (Quadro 1): a) capacidade funcional: itens 3a, 3b e 3j; b) aspectos físicos: item 4c; c) aspectos emocionais: item 5b; d) aspectos sociais: item 10; e) ambos os itens do domínio dor; f) saúde mental: itens 9b e 9f; g) vitalidade: item 9e.

O resultado é um sistema descritivo de saúde composto por 6 domínios chamado SF-6D, cujos itens estão demonstrados no Quadro 2. Cada domínio possui entre 2 e 6 níveis, e estes níveis possuem uma ordem em termos de gravidade da disfunção.

Estados de saúde

Um estado de saúde gerado por um sistema descritivo de um instrumento é obtido pela seleção de um nível de cada domínio, obedecendo à ordem dos domínios. Estes estados de saúde são denominados na literatura de estados “multiatributos”. Por exemplo, ao selecionarmos o nível 1 de cada domínio (p. ex. 111111), teremos o melhor estado de saúde gerado pelo instrumento, enquanto a seleção do último nível de cada domínio gera o pior estado (645655). Entre estes dois extremos, temos uma gama de possíveis combinações para formar os estados chamados intermediários. Exemplos de cada um são apresentados na Figura 1. O SF-6D gera 18.000 ($6 \times 4 \times 5 \times 6 \times 5 \times 5$) possíveis estados de saúde. Logisticamente, seria impossível realizar avaliação das preferências por todos estes estados. Então, uma amostra foi selecionada para realizar o procedimento de avaliação utilizando a técnica *standard gamble* para medida de preferências. Foram utilizados os mesmos estados de saúde avaliados no Reino Unido para permitir comparações, totalizando 247 estados intermediários. A seleção destes estados, feita previamente pela equipe da Universidade de Sheffield, foi realizada através de uma técnica ortogonal, que

permite selecionar randomicamente uma amostra a partir do conjunto dos 18.000 estados de saúde gerados pelo SF-6D. Esta técnica é aplicada através do procedimento Orthoplan contido no programa estatístico SPSS.¹

Amostragem

A etapa seguinte consiste na avaliação dos estados de saúde por um grupo de indivíduos. As diretrizes para estudos de custo-efetividade recomendam que QALYs sejam obtidos com preferências da comunidade em geral e não em grupos ou coortes de indivíduos com características específicas.²³ Portanto, uma amostra da população geral de Porto Alegre foi selecionada, considerando-se a composição por sexo e pelas faixas etárias: 20-29, 30-44 e 45-64 anos.

O cálculo de tamanho de amostra foi baseado no número de avaliações por estado, levando em conta que todos os estados de saúde deveriam ser avaliados por um número aproximadamente igual de indivíduos. Não há recomendação na literatura a respeito do número mínimo necessário de avaliações a fim de garantir consistência nos resultados. Existem evidências de que o impacto para o respondente de avaliar estados com múltiplos atributos aumenta ainda mais se o número de estados avaliados é grande.¹³ Considerando a experiência dos outros países que realizaram o mesmo protocolo, onde cada estado foi avaliado de 10 a 15 vezes em média e também levando em conta a disponibilidade em termos de recursos econômicos e tempo, estabelecemos um mínimo de 10 avaliações por estado. Cada indivíduo deveria avaliar 5 estados de saúde, sendo 5 intermediários além do pior estado. O cálculo resultou em 494 indivíduos.

Este foi um estudo transversal de base populacional. A seleção da amostra foi feita por um plano amostral complexo, por conglomerados. Estes foram

representados pelos setores censitários da cidade de Porto Alegre, segundo divisão feita pelo IBGE.² Com base na estimativa de adultos por domicílio, foram calculados o número de domicílios a serem visitados e o número de setores censitários. Foram aleatoriamente selecionados 108 setores censitários. Dentro de cada setor selecionou-se 7 domicílios de forma sistemática, convidando a participar todos os moradores que preenchessem os seguintes critérios de inclusão: 1) ter entre 20 e 64 anos de idade; 2) ser alfabetizado; 3) não ser portador de nenhuma doença física ou mental que impossibilitasse o preenchimento dos instrumentos.

Entrevistas e material utilizado

O protocolo de entrevistas e o material utilizado em cada etapa estão esquematizados na Figura 2.

Uma equipe de entrevistadores composta por estudantes de graduação de cursos da área da saúde recebeu treinamento intensivo para aplicação da técnica do *standard gamble*, que se realizou em 3 etapas: 1) Breve introdução para apresentação dos propósitos e da importância do estudo 2) Instruções sobre o uso do instrumento utilizando o material gráfico e o manual, com aplicação prática feita pelo coordenador do estudo ou coordenadores de campo através da simulação de uma entrevista com um dos membros da equipe; 3) Realização, pelo membro da equipe em treinamento, de uma ou 2 entrevistas com familiares ou amigos com aplicação do *standard gamble*. Este procedimento permitia identificar dúvidas e dificuldades na aplicação. Após os entrevistadores estarem familiarizados com a técnica do *standard gamble*, o treinamento para aplicação do restante do protocolo foi realizado e cada estudante realizava uma entrevista simulada inteira. A fim de garantir a qualidade dos dados, as primeiras entrevistas foram acompanhadas pelos

coordenadores de campo. Durante toda a duração da coleta de dados houve revisão do material trazido pelos entrevistadores logo após a chegada do campo.

As entrevistas foram realizadas no domicílio dos participantes e a apresentação da equipe ocorria através de uma carta onde estavam descritos o propósito do estudo, o tipo de tarefa que o entrevistado iria desenvolver e o tempo despendido na entrevista que foi estimado em torno de 40-60 minutos. Caso os moradores do domicílio não fossem encontrados em uma primeira visita, mais duas visitas em dias e horários diferentes, incluindo horário não comercial, foram realizadas, além de contato telefônico quando possível. Se após 3 visitas consecutivas não houve nenhuma possibilidade de contato com os moradores, foi considerado como perda. No planejamento da amostra, foi calculado um acréscimo de 10% sobre o número de setores censitários selecionados para reposição de perdas.

O protocolo iniciou com os entrevistados avaliando sua própria saúde através dos itens do SF-6D para que adquirissem familiaridade com a descrição de saúde que seria apresentada através dos estados de saúde.

A avaliação dos estados de saúde consistia primeiramente de um ordenamento de 8 estados, conforme as preferências do entrevistado, do melhor para o pior: 5 estados intermediários gerados pelo SF-6D, mais os estados extremos (o melhor estado, o pior estado e morte imediata). Para esta tarefa, foi apresentado um material gráfico, para que os entrevistados tivessem alguma referência para ordenar os estados conforme sua preferência. Foi utilizada uma escala numérica, onde 10 representava o estado mais preferido e 0 o menos preferido. Não era necessário alocar os estados em valores da escala, ela servia apenas como apoio. Os estados de saúde eram representados em cartões como mostra a Figura 1,

confeccionados na cor branca para evitar que as cores pudessem influenciar o ordenamento.

O próximo passo foi a aplicação do *standard gamble*, instrumento considerado “padrão” para medida de preferências por ser baseado em teorias econômicas.²¹ A técnica iniciou apresentando ao entrevistado uma descrição de um dos estados de saúde gerados pelo SF-6D. Após ler esta descrição, o entrevistado foi orientado a imaginar a situação hipotética na qual ele (ela) é confrontado com uma escolha. As opções disponíveis eram continuar a viver no estado descrito em um dos estados intermediários, ou optar por um *gamble* (ou “aposta”). Esta “aposta”, representada pela hipótese de realizar um tratamento, tinha 2 possíveis desfechos. O melhor desfecho seria o tratamento ter sucesso e ocorrer recuperação imediata da saúde perfeita, com uma probabilidade “ p ”. O pior desfecho seria falha no tratamento levando ao pior estado de saúde do SF-6D, com probabilidade “ $1-p$ ” de ocorrência. Estas probabilidades variavam de forma sistemática durante o exercício, até que o entrevistado fosse indiferente entre manter-se no mesmo estado intermediário ou aceitar a “aposta”. A representação esquemática do *standard gamble* pode ser vista na Figura 3. O material utilizado foi criado por uma equipe de pesquisadores da Universidade de McMaster, no Canadá¹³ e foi reproduzido aqui no Brasil com autorização dos autores. Os mesmos estados de saúde que o entrevistado havia ordenado na tarefa anterior eram utilizados para o *standard gamble*, agora representados por cartões coloridos. Os estados intermediários tinham a cor verde e ficavam em um bolso no local da “Escolha B” do “Quadro de chances”, mostrando 100% de chance de permanecer naquele estado. A “Escolha A” era a “aposta” que demonstrava do lado esquerdo a chance da intervenção funcionar e a pessoa recuperar a saúde plena e do lado direito as chances do desfecho oposto. Os

cartões do melhor e pior desfecho eram apresentados nas cores rosa e azul, respectivamente. As chances eram apresentadas nas mesmas cores dos cartões, em rosa e azul, em números acima dos cartões e também na forma de “gráfico de torta”. A figura 3 também mostra um exemplo de variação das chances. Estes passos eram repetidos para cada um dos 5 estados intermediários.

O fato de o *standard gamble* ter sido realizado tendo como pior desfecho o pior estado de saúde do SF-6D e não “morte imediata”, como é o padrão, exigiu mais uma etapa no processo. Para que os valores das preferências obedecessem à escala entre 0 (morte) e 1 (saúde perfeita), o pior estado de saúde também teve de ser avaliado para que os valores dos outros estados fossem ajustados de acordo com o valor dado ao pior estado. Então, havia uma sexta avaliação no *standard gamble*, que dependia de como o entrevistado tinha ordenado o pior estado, se melhor ou pior do que a morte: se considerou a morte pior, os desfechos da “aposta” eram saúde perfeita e morte e o pior estado era o cenário de certeza (“Escolha B”); se considerou a morte melhor do que o pior estado, os desfechos da “aposta” eram saúde plena e o pior estado e no cenário de certeza, morte imediata. Esta variação na técnica permite que estados considerados piores que a morte sejam avaliados.

Após o término do *standard gamble*, o entrevistador registrou o grau de dificuldade encontrado pelo entrevistado ao realizar o método.

A escolha dos 5 estados de saúde para cada entrevista foi randômica, mantendo-se fixos apenas os estados extremos, ou seja, o melhor e o pior estado e morte imediata.

A última etapa do protocolo consistia no preenchimento dos questionários SF-36 e WHOQOL-breve pelo entrevistado e a aplicação de um questionário sócio-demográfico pelo entrevistador.

Todos os participantes assinaram o Termo de Consentimento Informado aprovado pelo Comitê de Ética em Pesquisa do Hospital de Clínicas de Porto Alegre.

Algoritmo para o SF-6D

A partir dos valores das preferências obtidos pelo *standard gamble*, modelos de regressão foram construídos com os valores ajustados. A modelagem é feita considerando-se como desfecho (variável dependente) o valor atribuído ao estado de saúde pelo *standard gamble* e como preditores as 25 variáveis *dummies* criadas para representar cada nível de cada domínio do SF-6D. Os coeficientes representam a mudança do nível 1 de cada domínio, onde não há disfunção, para os outros níveis com graus progressivos de disfunção. Os valores dos coeficientes descrevem a magnitude desta mudança. Estes modelos são utilizados para construir a fórmula para cálculo dos escores gerados pelo sistema SF-6D. Os resultados da modelagem permitem assim que, a partir dos valores obtidos no SF-36 seja possível atribuir-se um valor de utilidade adaptado para esta população.

DISCUSSÃO

Este artigo teve como objetivo divulgar um método de pesquisa ainda pouco conhecido em nosso meio, para que os resultados gerados sejam mais bem compreendidos e mais facilmente aplicados. Instrumentos como o SF-36 e WHOQOL-breve já vem sendo utilizados no contexto nacional há alguns anos, mas os instrumentos baseados em preferências são de uso mais restrito devido à escassez de medidas disponíveis traduzidas e validadas para a população brasileira.

O principal resultado que está sendo gerado com os dados coletados através da metodologia descrita será um algoritmo para o SF-6D baseado em uma amostra

populacional de uma região do Brasil. Até então, pesquisadores interessados em utilizar índices de utilidade gerados do SF-36 necessitavam utilizar algoritmos desenvolvidos em outros países, não aconselháveis porque preferências podem diferir de uma cultura para outra. Ao término do estudo, o algoritmo gerado poderá ser empregado para estimativas de índices de utilidade. Estudos que tenham aplicado o SF-36 em populações específicas poderão obter valores de preferências, desde que os itens que compõem o sistema SF-6D (Quadro 2) tenham sido respondidos.

É necessário esclarecer que, das 3 etapas preconizadas para obtenção de instrumentos baseados em preferências citadas na introdução, os autores realizaram a segunda e a terceira. A primeira etapa, o sistema descritivo, é de autoria dos pesquisadores que disponibilizaram a versão brasileira do SF-36⁷, uma vez que o SF-6D é derivado deste. Apresentamos os itens do SF-6D aqui apenas para maior entendimento da origem dos estados de saúde.

O objetivo do projeto SF-6D original foi derivar índices de utilidade após o preenchimento do SF-36, possibilitando a obtenção de dois tipos de medida de saúde em uma única coleta. Porém, os autores do sistema descritivo SF-6D em língua portuguesa aplicaram o instrumento diretamente em uma população de pacientes com artrite reumatóide, calculando os escores a partir da fórmula britânica.⁴ Então, este procedimento também poderia ser feito utilizando o algoritmo gerado pelo presente estudo. Se a aplicação direta do SF-6D traria vantagens práticas em relação à aplicação do SF-36 ainda não está claro e poderia ser alvo de pesquisa futura.

Outro importante desfecho deste projeto será a avaliação das respostas dadas pelos indivíduos da amostra ao *standard gamble*, uma técnica complexa e

que, intuitivamente, poderia ser de difícil compreensão por uma parte da população brasileira com menor nível intelectual e menos acesso à informação.

Os dados deste estudo também serão utilizados para: a) comparação com resultados encontrados nos outros países que realizaram o mesmo protocolo, Portugal¹¹, Japão⁵ e China¹⁵, além dos dados britânicos, visando uma avaliação transcultural das preferências de indivíduos da população geral; b) descrição das médias dos escores dos instrumentos SF-36 e WHOQOL-breve em uma amostra da população geral, buscando uma “normalização”. Estes escores poderão ser utilizados como “padrão ouro” para comparação de médias obtidas em outros grupos específicos.

A importância da descrição desta metodologia reside na necessidade de disponibilizar um maior número de instrumentos baseados em preferências para realização de análises econômicas e conseqüente avaliação plenos de tecnologia para uso racional de recursos. Pesquisadores interessados neste campo de pesquisa poderão replicar os passos aqui delineados para validação de outros instrumentos no contexto brasileiro. Além disso, a melhor compreensão da metodologia destas medidas facilita e melhora a interpretação de seus resultados por parte dos usuários.

Referências bibliográficas

- (1) Statistical Package for Social Sciences. IBM Company 2010 March 6 Available from: URL: <http://www.spss.com/software/>
- (2) Censo Populacional 2000. Instituto Brasileiro de Geografia e Estatística Available from: URL: www.ibge.gov.br. Acessado em 03.03.10
- (3) Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol* 1998 Nov;51(11):1115-28.
- (4) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002 Mar;21(2):271-92.
- (5) Brazier JE, Fukuhara S, Roberts J, Kharroubi S, Yamamoto Y, Ikeda S, et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol* 2009 Dec;62(12):1323-31.
- (6) Campolina AG, Bortoluzzo AB, Ferraz MB, Ciconelli RM. Validity of the SF-6D index in Brazilian patients with rheumatoid arthritis. *Clinical and Experimental Rheumatology* 2009;27:64-71.
- (7) Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras Reumatol* 1999;39(3):143-50.
- (8) Drummond M.F., Sculpher M.J., Torrance G.W., O'Brien B., Stoddart G.L. *Methods for the Economic Evaluation of Health Care Programmes*. Third ed. Oxford: Oxford University Press; 2005.
- (9) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996 Aug 3;313(7052):275-83.
- (10) Feeny DH, George W, Torrance, William J, Furlong. Health Utilities Index. In: Bert Spilker, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Second ed. Philadelphia: Lippincott-Raven; 1996. p. 239-52.
- (11) Ferreira L.N., Ferreira P.L., Pereira L.N., Brazier J, Rowen D. A Portuguese Value Set for the SF-6D. *Value in Health*. In press 2010.
- (12) Froberg DG, Kane RL. Methodology for measuring health-state preferences--II: Scaling methods. *J Clin Epidemiol* 1989;42(5):459-71.
- (13) Furlong W, Feeny D, Torrance G, Barr R, Horsman J. *Guide to Design and Development of Health-State Utility Instrumentation*. McMaster University, editor. 1990. Hamilton, Ontário. Working Paper Series #90-0.

Ref Type: Serial (Book,Monograph)

- (14) Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993 Apr 15;118(8):622-9.
- (15) Lam CLK, Brazier J, McGhee SM. Valuation of the SF-6D Health States is feasible, Acceptable, reliable, and Valid in a Chinese Population. *Value in Health* 2009;11(2):295-303.
- (16) Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care* 2000 Sep;38(9 Suppl):II138-II150.
- (17) Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996 Oct 9;276(14):1172-7.
- (18) The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990 Dec;16(3):199-208.
- (19) Torrance GW. Toward a utility theory foundation for health status index models. *Health Serv Res* 1976;11(4):349-69.
- (20) Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002 Jun;11(4):341-53.
- (21) von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. First ed. Princeton: Princeton University Press; 1944.
- (22) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 Jun;30(6):473-83.
- (23) Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996 Oct 16;276(15):1253-8.

Quadro 1 - Estrutura do SF-36

DOMÍNIOS	ITENS
Capacidade Funcional	3a. Atividades vigorosas 3b. Atividades moderadas 3c. Levantar ou carregar mantimentos 3d. Subir vários lances de escada 3e. Subir um lance de escada 3f. Curvar-se, ajoelhar-se ou dobrar-se 3g. Andar mais de um quilômetro 3h. Andar vários quarteirões 3i. Andar um quarteirão 3j. Tomar banho ou vestir-se
Aspectos físicos	4a. Diminuir a quantidade de tempo 4b. Realizar menos tarefas 4c. Limitação em atividades 4d. Dificuldade no trabalho
Dor	7. Magnitude da dor 8. Interferência da dor
Estado geral de saúde	1. Avaliação global da saúde 11a. Adoecer mais facilmente 11b. Tão saudável quanto 11c. Saúde vai piorar 11d. Saúde excelente
Vitalidade	9a. Vigor/vontade/força 9e. Energia 9g. Esgotamento 9i. Cansaço
Aspectos sociais	6. Interferência na vida social 10. Interferência no tempo da vida social
Aspectos emocionais	5a. Diminuir quantidade de tempo 5b. Realizar menos tarefas 5c. Cuidado com atividades
Saúde mental	9b. Pessoa nervosa 9c. Deprimido 9d. Calmo/Tranquilo 9f. Desanimado/abatido 9h. Feliz

Quadro 2 - Itens do SF-6D

Capacidade funcional	
1	Sua saúde <u>não</u> dificulta que você faça atividades vigorosas
2	Sua saúde dificulta <u>um pouco</u> que você faça atividades vigorosas
3	Sua saúde dificulta <u>um pouco</u> que você faça atividades moderadas
4	Sua saúde dificulta <u>muito</u> que você faça atividades moderadas
5	Sua saúde dificulta <u>um pouco</u> para você tomar banho ou vestir-se.
6	Sua saúde dificulta <u>muito</u> para você tomar banho ou vestir-se
Aspectos físicos e emocionais	
1	Você <u>não</u> teve problemas com o seu trabalho ou alguma outra atividade diária regular como consequência de sua saúde física ou algum problema emocional
2	Você esteve limitado no seu tipo de trabalho ou em outras atividades como consequência de sua saúde física
3	Você realizou menos tarefas do que você gostaria como consequência de algum problema emocional
4	Você esteve limitado no seu tipo de trabalho ou em outras atividades como consequência de sua saúde física e realizou menos tarefas do que você gostaria como consequência de algum problema emocional
Aspectos sociais	
1	Sua saúde física ou problemas emocionais não interferiram com as suas atividades sociais em nenhuma parte do tempo
2	Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais em uma pequena parte do tempo
3	Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais em alguma parte do tempo
4	Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais na maior parte do tempo
5	Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais todo o tempo
Dor	
1	Você <u>não</u> teve <u>nenhuma</u> dor no corpo
2	Você teve dor, mas a dor <u>não</u> interferiu <u>de maneira alguma</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
3	Você teve dor, e a dor interferiu <u>um pouco</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
4	Você teve dor, e a dor interferiu <u>moderadamente</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
5	Você teve dor, e a dor interferiu <u>bastante</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
6	Você teve dor, e a dor interferiu <u>extremamente</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
Saúde Mental	
1	Você <u>não</u> tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>nunca</u>
2	Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida em <u>uma pequena parte do tempo</u>
3	Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>em alguma parte do tempo</u>
4	Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>na maior parte do tempo</u>
5	Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>todo o tempo</u>
Vitalidade	
1	Você tem se sentido com muita energia <u>todo o tempo</u>
2	Você tem se sentido com muita energia <u>a maior parte do tempo</u>
3	Você tem se sentido com muita energia <u>em alguma parte do tempo</u>
4	Você tem se sentido com muita energia em <u>uma pequena parte do tempo</u>
5	Você não tem se sentido com muita energia <u>nunca</u>

<p>Sua saúde <u>não</u> dificulta que você faça atividades vigorosas</p> <p>Você <u>não</u> teve problemas com o seu trabalho ou alguma outra atividade diária regular como consequência de sua saúde física ou algum problema emocional</p> <p>Sua saúde física ou problemas emocionais <u>não</u> interferiram com as suas atividades sociais em <u>nenhuma parte do tempo</u></p> <p>Você <u>não</u> teve <u>nenhuma</u> dor no corpo</p> <p>Você <u>não</u> tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>nunca</u></p> <p>Você tem se sentido com muita energia <u>todo o tempo</u></p> <p style="text-align: right;">EM</p>
<p>Sua saúde dificulta <u>muito</u> para você tomar banho ou vestir-se</p> <p>Você esteve limitado no seu tipo de trabalho ou em outras atividades como consequência de sua saúde física e realizou menos tarefas do que você gostaria como consequência de algum problema emocional</p> <p>Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais <u>todo o tempo</u></p> <p>Você teve dor, e a dor interferiu <u>extremamente</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)</p> <p>Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>todo o tempo</u></p> <p>Você <u>não</u> tem se sentido com muita energia <u>nunca</u></p> <p style="text-align: right;">EP</p>
<p>Sua saúde dificulta <u>um pouco</u> que você faça atividades moderadas</p> <p>Você realizou menos tarefas do que você gostaria como consequência de algum problema emocional</p> <p>Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais <u>em alguma parte do tempo</u></p> <p>Você teve dor, e a dor interferiu <u>um pouco</u> com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)</p> <p>Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida <u>em alguma parte do tempo</u></p> <p>Você tem se sentido com muita energia <u>em alguma parte do tempo</u></p> <p style="text-align: right;">E116</p>

Figura 1 - Representação dos cartões contendo estados de saúde. Na ordem de cima para baixo: melhor estado de saúde (EM), pior estado de saúde (EP) e um estado intermediário (E116).

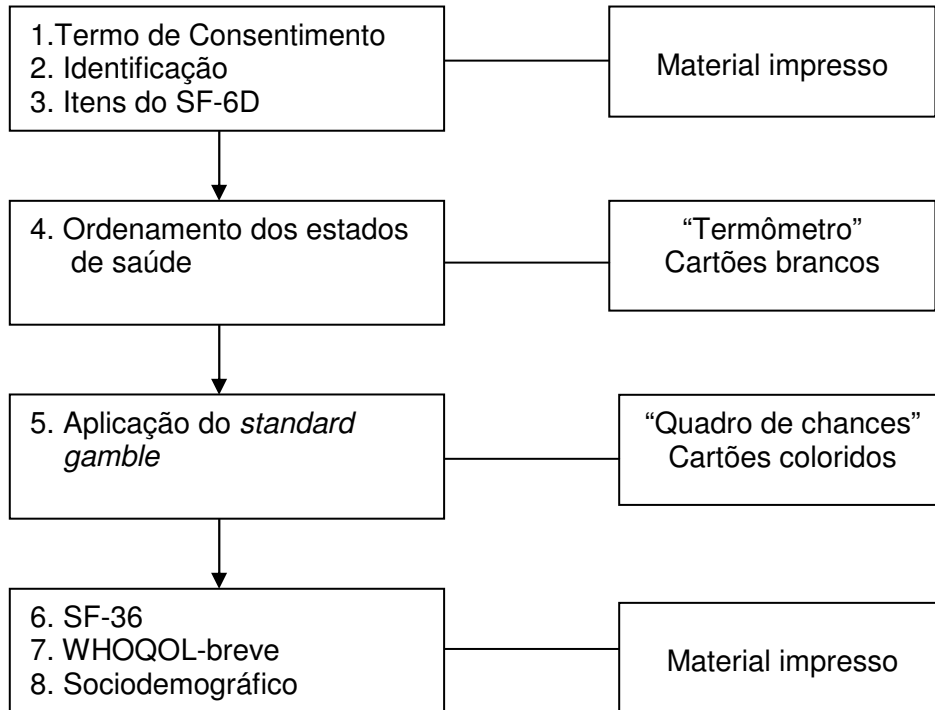


Figura 2 - Protocolo da coleta de dados

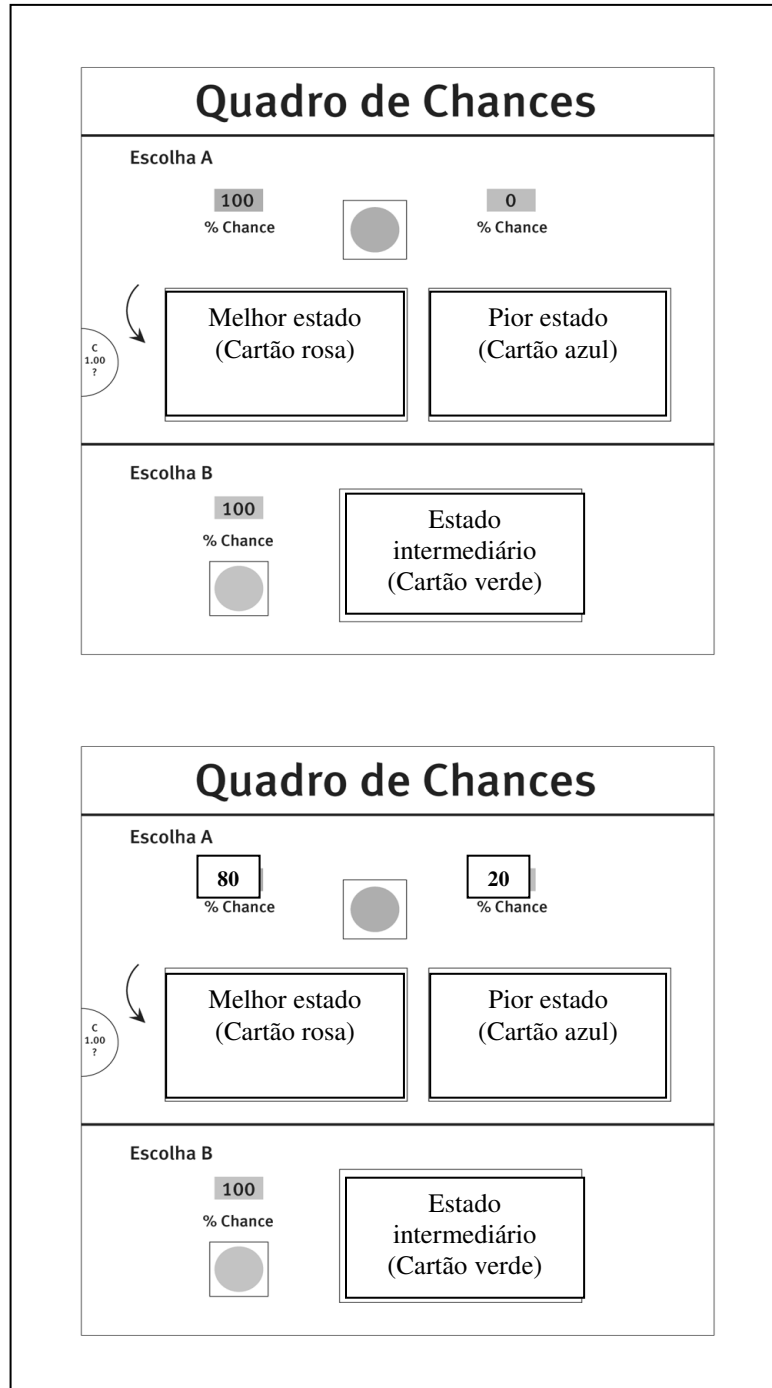


Figura 3 - Representação do material utilizado para realizar *standard gamble*, o “Quadro de Chances”. O quadro é composto por um círculo que gira manualmente e as chances são sistematicamente variadas e demonstradas nas janelas superiores da “Resposta A”.

ARTIGO 2

**INCONSISTENCIES AND SF-6D HEALTH STATES
VALUATION IN BRAZIL**

INCONSISTENCIES AND SF-6D HEALTH STATES VALUATION IN BRAZIL

Luciane N. Cruz ¹, Marcelo P. Fleck², Suzi A. Camey³, Juliana F. Hoffmann¹, Donna Rowen⁴, John E. Brazier⁴, Carisi A. Polanczyk^{1, 5}

Artigo a ser submetido ao Journal of Clinical Epidemiology

Graduate Studies Program in Epidemiology ¹, Graduate Program in Psychiatry², Statistics Department³, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; School of Health and Related research, University of Sheffield, UK⁴; Cardiology Division of Hospital de Clinicas de Porto Alegre⁵, Brazil.

Corresponding author:

Luciane Nascimento Cruz

Rua Ramiro Barcelos, 2350/second floor

Porto Alegre – RS – Zip Code 90035-003

Phone: 55 51 3356-7880

Fax: 55 51 3342-7959

E-mail: Incruz@terra.com.br

ABSTRACT

Objective: To assess socio-demographic factors associated with inconsistent responses in the valuation for health states generated by SF-6D. **Study Design and Setting:** Data from the SF-6D valuation survey were used. A population-based sample of 528 Southern Brazilian adults valued 247 SF-6D health states using Standard Gamble (SG) method. Each participant evaluated a group of 5 health states and the worst state generated by the instrument. Utility values were considered inconsistent when they did not vary in a range of different health states and when the value for the worst state was illogical. **Results:** 58 subjects (11%) gave the same value for all 5 health states presented to them and 88 (17%) valued the worst state illogically. The group with inconsistent responses were older ($p=0.018$), less educated ($\chi^2=19.90$ $p<0.01$) and from a lower income group ($\chi^2=16.09$ $p=0.007$). Invariance was associated with age and a shorter time spent on standard gamble with prevalence ratios of 1.7 (CI 95% 1.1-3.4) and 2.6 (CI 95% 1.8-3.7), respectively. **Conclusion:** Inconsistencies in utility ratings for multi-attribute states are frequent, mainly in some groups as older and less educated subjects. In future research, we need to determine if these inconsistencies would affect our final SF-6D index.

Keywords: inconsistency; health state valuation; standard gamble; SF-6D; utility.

RESUMO

Objetivos: Avaliar os fatores sociodemográficos associados com respostas inconsistentes na avaliação dos estados de saúde gerados pelo SF-6D. **Métodos:** Os dados do inquérito de avaliação dos estados de saúde do SF-6D foram utilizados. Uma amostra de 528 adultos da população geral de Porto Alegre avaliou 247 estados de saúde gerados pelo SF-6D utilizando o método *standard gamble* (SG). Cada participante avaliou um grupo de 5 estados de saúde além do pior estado gerado pelo instrumento. Valores de utilidade foram considerados inconsistentes quando diferentes estados de saúde tiveram o mesmo valor e quando o valor atribuído ao pior estado foi ilógico. Resultados: 58 indivíduos (11%) atribuíram o mesmo valor para todos os 5 estados e 88 (17%) avaliaram o pior estado de forma ilógica. O grupo com respostas incoerentes era mais velho ($p = 0,018$), com menor escolaridade ($\chi^2 = 19,90$ $p < 0,01$) e renda familiar mais baixa ($\chi^2 = 16,09$ $p = 0,007$). Invariância nos valores foi associada com idade e um menor tempo gasto no *standard gamble*, com razões de prevalência de 1,7 (CI 95% 1,1-3,4) e 2,6 (CI 95% 1,8-3,7), respectivamente. Conclusão: Inconsistências na avaliação de estados de saúde com múltiplos domínios são frequentes, principalmente em indivíduos mais velhos e com menor nível de escolaridade. Mais estudos serão necessários para determinar se valores inconsistentes afetariam o modelo final para construção do algoritmo SF-6D.

Palavras-chave: Inconsistências; avaliação de estado de saúde; *standard gamble*; índices de utilidade

INTRODUCTION

The use of preference-based measures of quality of life for clinical decision-making and economic analyses has increased exponentially in the last decade. The expanded development of these instruments has been occurring mainly due to the fact that they generate QALYs (Quality Adjusted Life Years), an outcome measure recommended by several regulatory agencies such as the National Institute for Clinical Excellence (NICE) (1) in the United Kingdom and the Panel on Cost-effectiveness in Health and Medicine (2) of the US Public Health Service for use in cost-effectiveness studies.

An important step involved in the development of these measurements is the determination of numerical values that represent the population's preferences for those health states generated by the instrument (3). These values can be obtained through methods such as the visual analog scales, time trade-off (TTO) or standard gamble (SG), (4) and are used for the construction of a model, or algorithm, that can predict the values for all possible health states generated by the descriptive system, once it is logistically impossible to conduct an assessment of all possible states.

The growing application of these instruments has led researchers to find out that the preference values per health states do not always conform to a previously established logic, generating what has been called in literature as "inconsistency"(5). There is no consensus related to its concept, being defined by each author depending on the context of the study. In general, a logical inconsistency occurs when a health state is recognizably better compared to another, but receives a lower value from the evaluator (6). Other authors also considered as an inconsistent response the invariance, which occurs when the individual attributes the same value to different health states (7). Badia et al. (8) conceived as "criterion inconsistency" when different methods (ranking, VAS, TTO) produce a disagreeing ordering of preferences.

Despite the lack of a definition criterion, there is a common concern in understanding the factors associated to the inconsistencies. Inconsistent values can be associated with both intrinsic characteristics of the subject and methods to

measure preferences, and several parameters have been related to them such as age (5;9) level of education (5;7) and race (6).

Nonetheless, there are scarce studies reporting inconsistencies considering the wide use of these instruments. Data related to the inconsistencies found in valuations of health states based on EuroQol-5D has been that most prevalent in literature, and reports for the Health Utility Index Mark 2 were also recorded (10). The rates of inconsistencies have varied, depending on the concept and form of evaluation, from 5% to 89% in the various studies (7;8;10;11)

Inconsistencies found in valuations of SF-6D health states are lacking in literature, since it is a more recent instrument. The countries that conducted valuation studies of the SF-6D health states (12;13) replicated the protocol of the original British study (14), excluding inconsistent values from the final index, but an association between inconsistency and characteristics of the sample was not presented. In a similar process of cultural adaptation of SF-6D for a Southern Brazilian population, a survey to value SF-6D health states was conducted. The main results will be described in another paper. The objective of this study was to evaluate the subpopulation of individuals that was excluded from our final model due to invalid values, seeking to contribute with a body of evidence directed toward understanding the factors related to the inconsistencies in the preference elicitation process.

METHODS

The study followed the same protocol of the SF-6D original British study(14).

Valuation of health states

With the aim of deriving the descriptive system of Brazilian SF-6D, the authors used the same derivation methodology from the SF-36 items used in the original study of the United Kingdom (14). For this, the version of the SF-36 translated and validated in Brazil by Ciconelli et al. (15) was used. The result is a six-dimensional health state classification called SF-6D. From this descriptive system, health states were generated to be valued by the respondents. The same health states selected in

the British protocol were used, in order to make comparisons, totaling 247 intermediate states. These were evaluated by the population sample using the standard gamble technique applied through graphical material, the “chance board”. Each individual evaluated 5 states plus the worst state generated by the instrument. The health states were presented in the form of colored cards. Before applying the standard gamble, the individuals were invited to rank the health states by order of preference, from the best to the worst, including, in addition to the 6 states referred to above, the best state generated by the instrument and immediate death. In order to ensure that the respondents did not have to rank very similar health states, the 247 states were divided into 5 subgroups according to the level of severity. This criterion was evaluated through a score based on the sum of scores of the levels of each dimension present in each state. One state from each subgroup was selected to present to each respondent.

After the ranking exercise, the standard gamble was applied, in which two alternatives were proposed to the participant: a certain prospect of living in an intermediate state defined by the SF-6D and the uncertain prospect of two possible outcomes, the best state described by the SF-6D or the worst state. The probability of the best outcome occurring was varied until the respondent was indifferent between the certain and the uncertain prospects. The probabilities varied systematically in a “ping-pong” with props version of standard gamble (16). This technique is designed to make the interview as straightforward as possible, by leading the interviewer through a group of questions depending as the respondents answer to the previous question and minimize the risk of interviewer variation.

At the end of the standard gamble, the interviewer asked the participant about the level of difficulty perceived on performing the task, the answers being standardized in a Likert scale: “very easy”, “easy”, “somewhat difficult”, “difficult” and “very difficult”.

In the study protocol, a modified version of the standard gamble was used, in which the “immediate death” outcome is replaced in the lottery by “worst health state”. Thus, there is need to adjust the values of the intermediary states according to the value given by each subject to the worst health state. Taking “ P ” to be the value given to the worst state and “ SG ” the values of the intermediate values, the formula would be: $SGADJ = SG + (1-SG) P$. “ $SGADJ$ ” represents the standard gamble

adjusted score. This converts the values such that they are in a scale between 0 and 1, which are the reference values for utility indices. Evaluation of the worst health state was done after evaluating the fifth card and definition of the anchor points of the gamble at this point depended on how the respondent had ordered the worst state, if better or worse than death, in the ranking exercise: 1) if better than death, the respondent must decide between the certain prospect of being in the worst state and the uncertain prospect of full health or immediate death; 2) if worse than death, the choose was between the certain prospect of death and the uncertain prospect of full health or the worst state.

After obtaining the utility values for the health states, the next step was the construction of a predictive model for the values of all possible health states generated by the SF-6D. We used the same rules as the British study to select the values that would be considered valid and would enter the model. The exclusion criteria were: a) individuals without a valid value for the worst state, avoiding to adjust the values of all other intermediary states; b) not have valuated two or more health states; c) have given the same value to the 5 states evaluated.

Sampling

The sample consisted of individuals selected from the general population of Porto Alegre, a capital city in the South of Brazil. The sample size was calculated such that each health state was evaluated at least 10 times. Since each individual would evaluate 5 intermediary states, the sample size estimation was 494 individuals.

A sample of the general population aged 20 to 64 years was selected. A two-stage cluster random selection design was used. These primary units were represented by census sectors of the municipality according to division made by IBGE (Brazilian Geography & Statistics Institute) (17). In a second stage, households were systematically selected in each census sector. To calculate the number of households to be visited, the average number of adults per household was considered, and the population's proportion in each one of the strata the study aimed to reach, that is, men and women in the age ranges of 20-29, 30-44, 45-64 years. The interviewers visited 7 domiciles in each of the 108 selected census sector,

inviting all dwellers who met the following inclusion criteria to take part in the study: have age between 20 and 64 years, be literate and not have any physical or mental incapacity that would prevent reading and understanding of the instruments.

Interview protocol

A team of interviewers made up of graduation students from health-related courses received intensive training to apply the standard gamble technique and the rest of the protocol. The interviews were conducted in the respondents' own home and the team was presented by means of a letter that described the purpose of the study, the type of task that the interviewer would develop and the time spent in the interview, which was estimated at about 40-60 minutes. The protocol consisted of the following instruments, applied strictly in this order: 1. identification of the participant; 2. application of the SF-6D instrument for the participant to become acquainted with the health description that would be presented in the cards containing the health states; 3. ranking exercise; 4. standard gamble application; 5. verification of the level of difficulty found in SG task; 6. application of the instruments SF-36 and WHOQOL-brief ; 5. socio-demographic questionnaire.

Before the interview, all participants signed the Informed Consent Form approved by the University's Institutional Review Board.

The study described in this paper is part of a broader project that had as objectives, in addition to valuation of the SF-6D health states, standardization of scores of the SF-36 and WHOQOL-brief instruments in the general population.

Variables

For data analysis, the sample was divided into two groups of individuals: the group who presented inconsistencies in health states valuation and another group with valid values (consistent respondents). Inconsistencies were classified according two criteria: a) the respondent gave the same value for the 5 health states. This kind of inconsistency was called invariance; b) the worst health state value was invalid, rendering it impossible to calculate the value adjusted for the other states. Invalid values represented one of two situations: 1) the individual preferred to remain in the

worst health state for the rest of his (her) life to accept the gamble that offered 100% chance of returning to full health or; 2) preference for the worst health state was different in the ranking task of ordering and in the standard gamble. For example, in the ranking, the interviewee considered the worst state as better than death, but in the Standard Gamble, preferred to accept the gamble having 100% chance of dying. Since this response violates the expected logic, we decided referring to it herein as logical inconsistency, being different from invariance. This type of inconsistency could be valid for any health state besides the worst state, but the original protocol sought to be more conservative in relation to the exclusions of individuals with only some inconsistent values, keeping them in the model.

Once the study's objective was to evaluate socio-demographic factors related to the inconsistent responses, these two groups were compared in relation to the following variables: gender, age, race, marital status, religiousness, level of education (in years of study), job status, economic class, sufferer of a disease or not (self-report), level of difficulty reported and time spent in performing the standard gamble exercise.

The economic class was evaluated through an index called Brazil Criterion that classified the population into 7 classes according to its purchasing power and level of education of the head of the family (18). The division and its equivalence in terms of monthly family income in American dollars are, approximately: Class A1: US\$ 3, 800; Class A2: US\$ 2,300; Class B1: US\$ 1,400; Class B2: US\$ 800; Class C: US\$ 460; Class D: US\$ 212; & Class E: US\$ 103.

Some variables were classified as follows and included in regression models: race – white (reference category) and non-white; marital status – married (reference category) and not married; economic class – A (reference category), B & C; years of study – up to 8 years, from 8 to 11 years and 12 years or more (reference category); job status – active (reference category) or inactive, age – up to 44 (reference category) and 45 or more years and time spent on performing the standard gamble – up to 11 minutes and 12 minutes or more (reference category).

Statistical analyses

Continuous data are expressed as means (standard deviation) and categorical

in percentage. Pearson's chi-square test was used to verify the association of socio-demographic characteristics with two groups of individuals. The differences between means were compared by the Student's *t* test or Mann-Whitney test for not normally distributed variables.

A Poisson regression model with robust variance was built to examine the relationship of the inconsistent responses with the socio-demographic variables and time spent in conducting the standard gamble. The results are presented through the prevalence ratios among the group that had both types of inconsistent responses and the group with valid responses according to each one of the above-mentioned independent variables. The dependent variable was the type of response to the standard gamble, if consistent or not. After constructing the univariate models of the dependent variable (response type – if consistent or not), with the variables described in the Methods section, those presenting *p* value < 0.20 were selected for the final model. These variables were: age, economic class, education (years of study), job status, presence of disease (self-report) and time spent in the standard gamble task. In the multivariate model, the variables that maintained statistical significance were economic class, time spent in the standard gamble task and age. The level of education was closer to significance in relation to the other variables. Considering that this is a factor intuitively associated with greater difficulty of understanding a technique like standard gamble, we decided to examine the relation of these variables through an analysis of correspondence, which describes the relation between two nominal variables and also the relations between their categories, represented by the proximity between two points on the graph.

For all tests a significance level of 0.05 was established. The analyses were conducted using the SPSS Program version 16.0 (SPSS Inc., Chicago).

RESULTS

Study population

A total of 889 households were visited from July 2007 to March 2008. It was possible to conduct interviews in 354 (40%) of them, because for the others it was impossible to contact dwellers after 3 consecutive visits or people refused to

welcome the study team. Out of 846 eligible individuals identified, 537 accepted to take part, obtaining a response rate of 64%, similar to that of 65% found in the population of the British study. The number of eligible individuals includes all the households in which the research team was able to contact, even households where people refused to participate, because we could collect information regarding the number of people aged from 20 to 64 in the contacted addresses.

Out of 537 individuals, 9 (1.6%) were excluded because they did not complete the ranking exercise or the standard gamble. The final sample consisted of 528 individuals. 146 (28%) participants were classified as inconsistent respondents according to the criteria described in the methods section and detailed below.

All the respondents evaluated the worst health state, but in 58 (11%) the value was considered invalid according to criteria described above. There were 88 individuals (17%) whose value for the 5 health states did not vary, also being excluded. The results below describe the analysis of comparison between these subjects with invalid values (excluded) and included subjects with valid responses.

Table 1 shows the sample distribution in relation to the age, gender and level of education compared to the general population. Overall, the sample studied represents the general population regarding sex and age, with under representativeness only in the male gender in the ages between 30 and 44 years and female gender between 20 and 29 years. The mean years of study was greater in the study population, mainly in individuals aged over 60 years, but we must point out that the data available for the general population also includes individuals over 64 years.

Comparison of the socio-demographic profile between the groups is represented in Table 2. The variables with statistical significance were age, economic class and years of study. The subjects with inconsistent responses had 45 years of age or more, from the lowest economic class and with less years of study.

Interviews

Through the chi-square statistics, there was no significant difference in the level of difficulty in performing the standard gamble exercise between the groups. However, the proportion of individuals who considered the task “very easy” was greater in the group with inconsistent responses, while the percentage of participants

who found the exercise “hard” was greater in interviewees with coherent responses (Figure 1).

The mean time spent to conduct the standard gamble was 13 (11) minutes in the group with invariant responses, 14 (8) minutes in the group with invalid value for the worst state and 16 (8) minutes in the group with consistent responses. Only between the invariant and consistent group was there significant difference ($p = 0.006$)

Regression analyses

The results of multivariate model (model 1) can be seen through Figure 2. Class C presented a prevalence ratio of inconsistent responses (PR) of 1.7 (CI 95% 1.0-3.1) compared to Class A and individuals between 45 and 64 years had PR of 1.4 (CI 95% 1.0-2.1). In relation to the time spent in the standard gamble task, the interviewees that spent up to 11 minutes presented double prevalence of inconsistent responses in relation to the group that spent 12 minutes or more in the exercise (CI 95% 1.6-2.6).

In order to examine if the associations between inconsistent response and the predictive variables were similar depending on the type of inconsistency, if invariance or logic, two other models were constructed. One of them had as dependent variable the invariant response and the other logical inconsistency, and the results are represented in Figures 3 and 4, respectively. The model with invariant responses (model 2), only the time spent in the SG and age variables showed statistical significance, with higher prevalence ratios in relation to those obtained in the general model. Individuals whose time used to perform the SG was up to 11 minutes had prevalence of invariance 2.6 (CI 95% 1.8-3.7) times the prevalence of participants who answered in 12 minutes or more. The older individuals, aged 45-64 years, had prevalence of invariant responses 1.7 (CI 95% 1.1-3.4) times the prevalence of individuals aged up to 44 years. Figure 4 shows that no variable maintained statistical significance in the model with logical inconsistency, since all the confidence intervals crossed the unit. An inconsistent response given to the worst state does not seem to be explained by any of the variables considered in our analyses.

The results regarding to correspondence analysis are displayed in Figure 5,

which shows that there seems to be a relation between higher level of education and consistent responses. The points representing inconsistent responses are more spread out, but are close to the point of the category with lowest level of education.

DISCUSSION

The results of this study confirm findings in literature in relation to factors related to inconsistencies in the measurements of utility values, but with the innovation of evaluating this relation using health states of an instrument for which these data were still not available. To our knowledge, this is the first study that evaluated characteristics of the individuals associated with invariance and logical inconsistency in the values of health states preferences generated by SF-6D.

The percentage of individuals with invariant values among the states evaluated was higher in our population (17%) in relation to the British population (10%) (14) and other countries that also conducted the same protocol, China (2%) (12) and Japan (12%) (13). On the other hand, the number of individuals who failed to attribute valid values to the worst state (11%) was equal to the number found by the Chinese and Japanese (11%) and lower than that of the United Kingdom (15%). The greater percentage of invariance found in our population could be associated with cultural and not just demographic differences, since the type of population selected for the studies was uniform among all, that is, samples generated from the general population. These data raise an interesting question for future research that of conducting a comparative transcultural study seeking a comprehensive evaluation of inconsistencies related to socio-cultural factors. The fact that a standard protocol was used to gather data in these countries allows greater exploration of factors related to the respondent, since variables related to the technique used to obtain preferences, the form of presenting the health states and the states themselves did not vary.

Studies reporting invariance as an inconsistency are less common in literature, even for other instruments. Bravata (7) found 24% of invariance in a population made up of adult patients above 65 years of age. The most common type of inconsistency studied is that of violation in the logical ordering of the health states. EuroQoL-5D, for example, has 5 dimensions with 3 levels each. It is expected that state 21121, for

example, receives a value greater than state 32223, since in the latter all dimensions have their levels in category with higher dysfunction in relation to the former state. When this does not occur, it is regarded that the respondent had a response with logical inconsistency (11). However, it is more difficult to assume that SF-6D has any type of ordering among all the states, since there is a higher number of levels in each dimension (from 4 to 6) and 6 dimensions, being possible to generate 18,000 health states. With this range of possible combinations, there will be states with higher level in physical attributes and lower in psychological attributes, for example, and differences in values will not mean inconsistencies, but rather the individual's preference for a certain dimension. Therefore, this type of "inconsistency of order" was not evaluated in our study.

Intrinsic factors related to the individual were associated with inconsistent responses. Age proved to be a variable related to invariance, when individuals aged over 45 years had almost double prevalence of invariant responses compared to the group aged less than 45 years. This relation between older age and inconsistencies has also been found in several other studies (5;6;8). The reasons for this finding are not clear. Perhaps, the ability to maintain attention and interest in a task with ludic nature such as the "chance board" technique can be implied in the relation between age and inconsistency. The nature of this association can be better studied in qualitative studies, in which the factors related to the experience of eliciting health state preferences can be better explored.

Individuals from lower economic class also had a higher prevalence of inconsistent responses. In our population, a good economic condition guarantees greater access to information, health care and even a greater social insertion. Consequently, the more well-off classes have greater opportunity of attaining conditions required to make decisions about health, conditions that go beyond the knowledge obtained through school. The study data have shown a relationship between more years of study and consistent responses, as well as a greater proportion of people with over 12 years of study in the group with consistent responses in the chi-square test. Other authors also reported this finding (5;6-8), which confirms previous expectations that less instructed people found difficulty in interpreting a health state with multiple attributes and in understanding the measuring technique (3).

The group with inconsistent responses had greater proportion of individuals who regarded the standard gamble task as "very easy" and "easy". These data have already been reported by New Zealand authors (6) and we agree author's interpretation for this finding that the respondents did not put in much effort during the standard gamble task or did not understand the nature of the technique. The people may have answered the questions at random, without considering that they could make rational choices based on explicit probabilities of improving or worsening their health. The fact of putting in less effort to perform the standard gamble could also explain the less time spent in this task observed in individuals who gave the same value for all the health states.

The importance of identifying factors of the individuals related to inconsistencies resides in the need for answers to some questions that represent a challenge to researchers from this area: 1) If the socio-demographic factors are implied, would exclusion of individuals with inconsistent responses lead to an under-representation of some population group in the final model? 2) On the other hand, would inclusion of inconsistent values in the model compromise its predictive capacity and consequently generation of biased utility values for use in economic analyses? Our study contributed toward answering the first question in part, showing that individuals with inconsistent values form a group of greater age, of a lower economic class and with lower level of education. Some authors defend the strategy of trying to reduce the rate of inconsistencies in these subgroups through actions such as, for example, allowing these people to practice the technique more before answering the questions or pointing out the inconsistencies and allowing them to review their inconsistent answers (8). The problem of this approach is that it violates the standardization required in this type of interview.

We should acknowledge some caveats in our study. We do not know whether inclusion of individuals with inconsistent responses in the model would modify its parameters. These exclusions were a trade-off between the decision to keep the estimated sample size and the construction of models containing only values considered consistent. Furthermore, individuals considered "inconsistent respondents" could just be people who have a certain risk attitude, as, for example, people who gave the same values for different health states.

While we recognize some caveats of the study, it is important to emphasize the difficulty of conducting a population survey in our country. Due to high rates of urban violence in our city, part of the population lives in buildings with security systems that greatly hinder access to residents. For these reasons, it was necessary to adopt the strategy of replacement of losses and refusals, visiting a number of households larger than planned in order to obtain the required number of interviews.

In conclusion, our results suggest that some population subgroups may not produce consistent results for utility values measured through standard gamble. This must be considered by the researchers when choosing the instruments to be used to measure preferences. On the other hand, these findings reinforce the importance of instruments based on preferences, such as SF-6D, which generate utility values without the need to directly apply complex techniques like the *standard gamble*.

Acknowledgements: We thank the Research Incentive and Event Fund of Hospital de Clínicas de Porto Alegre for the financial aid in translating this article. Dr. Luciane Cruz received graduate research scholarship from CAPES, Brazil. Prof. Polanczyk and Prof. Fleck received a research scholarship from CNPq/Brazil

This study was funded by CNPQ/Brazil (Edital MCT-CNPq / MS-SCTIE-DECIT – Nº 36/2005).

References

- (1) National Institute for Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal. London: National Health Services; 2004.
- (2) Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996 Oct 16;276(15):1253-8.
- (3) Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. *J Health Econ* 1996 Apr;15(2):209-31.
- (4) Froberg DG, Kane RL. Methodology for measuring health-state preferences--II: Scaling methods. *J Clin Epidemiol* 1989;42(5):459-71.
- (5) Dolan P, Kind P. Inconsistency and health state valuations. *Soc Sci Med* 1996 Feb;42(4):609-15.
- (6) Devlin NJ, Hansen P, Kind P, Williams A. Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Econ* 2003 Jul;12(7):529-44.
- (7) Bravata DM, Nelson LM, Garber AM, Goldstein MK. Invariance and inconsistency in utility ratings. *Med Decis Making* 2005 Mar;25(2):158-67.
- (8) Badia X, Roset M, Herdman M. Inconsistent responses in three preference-elicitation methods for health states. *Soc Sci Med* 1999 Oct;49(7):943-50.
- (9) Froberg DG, Kane RL. Methodology for measuring health-state preferences--III: Population and context effects. *J Clin Epidemiol* 1989;42(6):585-92.
- (10) Wang Q, Furlong W, Feeny D, Torrance G, Barr R. How robust is the Health Utilities Index Mark 2 utility function? *Med Decis Making* 2002 Jul;22(4):350-8.
- (11) Lamers LM, Stalmeier PF, Krabbe PF, Busschbach JJ. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making* 2006 Mar;26(2):173-81.
- (12) Lam CLK, Brazier J, McGhee SM. Valuation of the SF-6D Health States is feasible, Acceptable, reliable, and Valid in a Chinese Population. *Value in Health* 2009;11(2):295-303.

- (13) Brazier JE, Fukuhara S, Roberts J, Kharroubi S, Yamamoto Y, Ikeda S, et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol* 2009 Dec;62(12):1323-31.
- (14) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002 Mar;21(2):271-92.
- (15) Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras Reumatol* 1999;39(3):143-50.
- (16) Furlong W, Feeny D, Torrance G, Barr R, Horsman J. *Guide to Design and Development of Health-State Utility Instrumentation*. [Paper #90-9]. 1990. Canada, Centre for Health Economic and Policy Analysis. McMaster University. Working Paper Series. Ref Type: Serial (Book, Monograph).
- (17) Censo Populacional 2000. Instituto Brasileiro de Geografia e Estatística Available from: URL: www.ibge.gov.br. Accessed 04.04.10
- (18) Critério Brasil 2003. Associação Nacional de Empresas de Pesquisa. Available from: URL: www.anep.org.br. Accessed 07.03.2010

Table 1 - Socio-demographic comparison between total sample and the general population of Porto Alegre

	Sample total (n=528) %	General Population * %
Men		
20-29 years	14	15
30-44 years	10	16
45-64 years	16	15
Women		
20-29 years	12	15
30-44 years	19	19
45-64 years	29	20
Economic class **		
A1	1	1
A2	13	5
B1	17	7
B2	24	17
C	38	38
D	7	28
E		5
Years of study according to age range* (mean) †		
20-24	11.8	10.1
25-59	11.7	8.8
60 or more	10.6	5.8

*Source: IBGE: www.ibge.org.br ** Source: Brazil Criterion - www.abep.org

† Age ranges modified for comparison with data available for the general population

Table 2 - Socio-demographic characteristics between included and excluded groups

Variables	Included (n=382) %	Excluded (n=146) %	<i>p</i>
Age (mean)	40.6	43.6	0.02
Gender			
Male	41	37	0.36
Female	59	63	
Race			
White	82	79	0.54
Black	9	12	
Others	9	9	
Religion			
Yes	51	61	0.04
No	49	39	
Marital status			
Married	63	60	0.42
Not married	37	40	
Economic class			
A1	1	2	0.007
A2	15	7	
B1	19	14	
B2	25	21	
C	34	48	
D	6	8	
Years of study			< 0.001
Up to 4 years of study	4	8	
5 to 8 years	16	31	
9 to 11 years	35	32	
12 years or more	45	29	
Job status			
Employed	54	44	0.28
Informal job	8	9	
Unemployed	7	5	
Housewife	14	21	
Student	4	6	
Retired	7	10	
Other	5	5	
Presence of disease			
Yes	46	53	0.14
No	54	47	

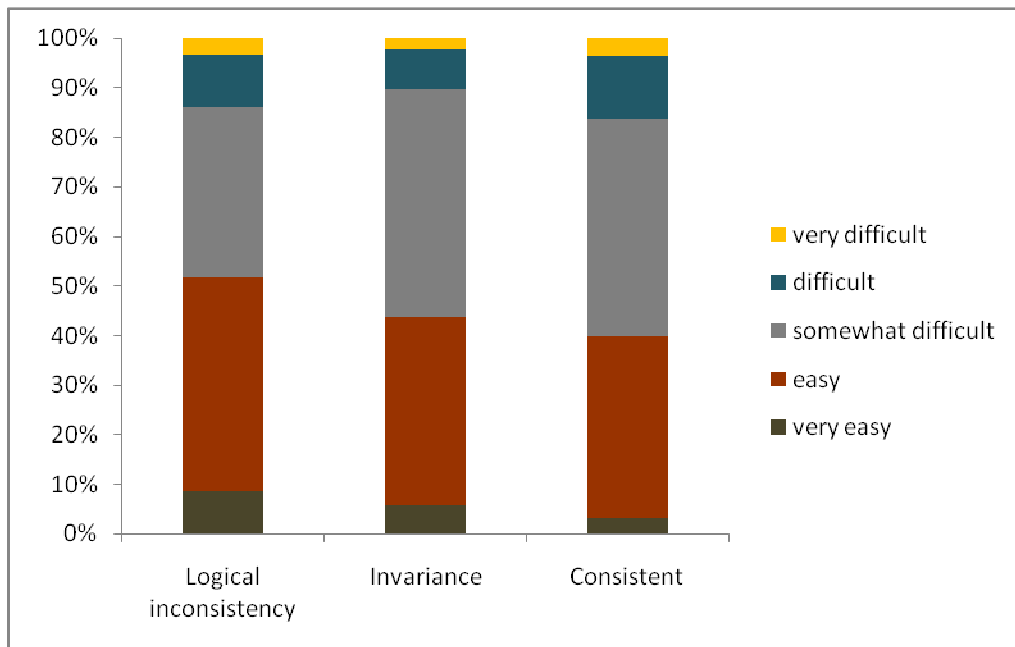


Figure 1 - Level of difficulty found on performing the Standard Gamble in the 3 groups, regardless of the response type ($\chi^2=7.7$ $p = 0.462$)

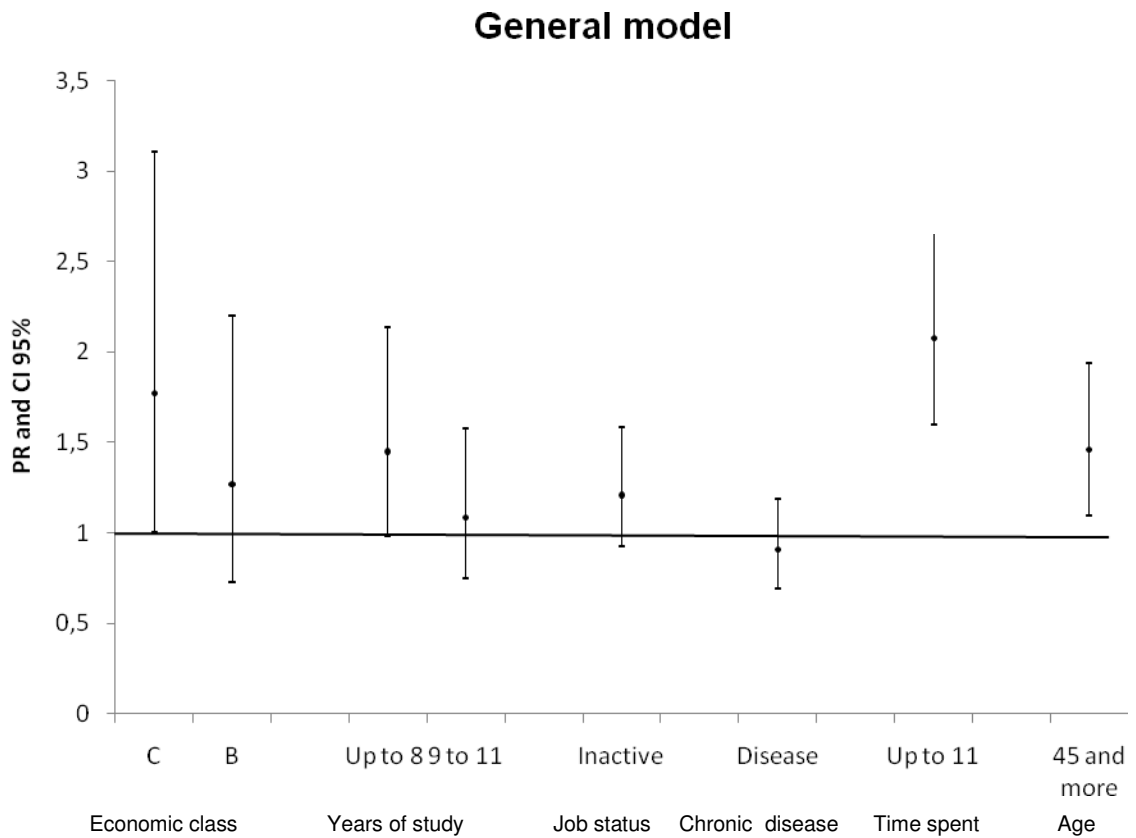


Figure 2 - Representation of regression model 1, showing Prevalence Ratios and CI 95% among inconsistent and consistent responses according to the following independent variables (from left to right): economic class, years of study, job status, presence or not of disease, time spent in the standard gamble (in minutes) and age.

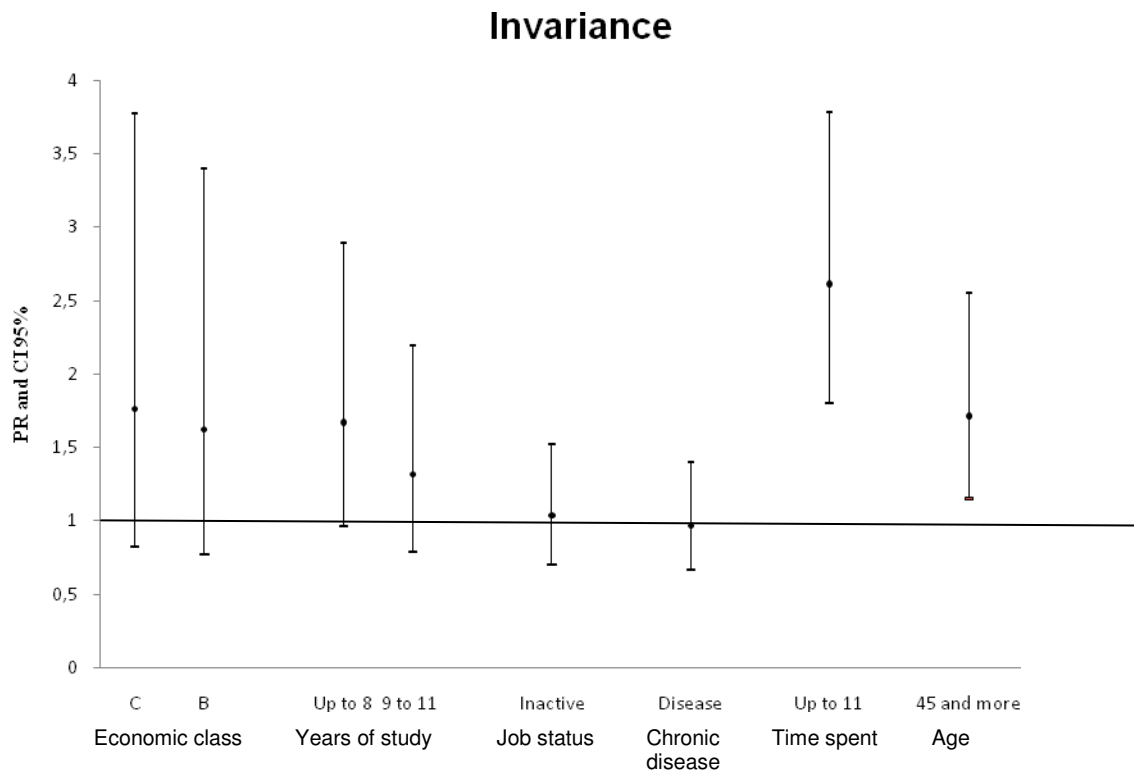


Figure 3 - Representation of regression model 2, showing Prevalence Ratios and CI 95% among invariance and consistent responses according to the following independent variables (from left to right): economic class, years of study, job status, presence or not of disease, time spent in the standard gamble (in minutes) and age.

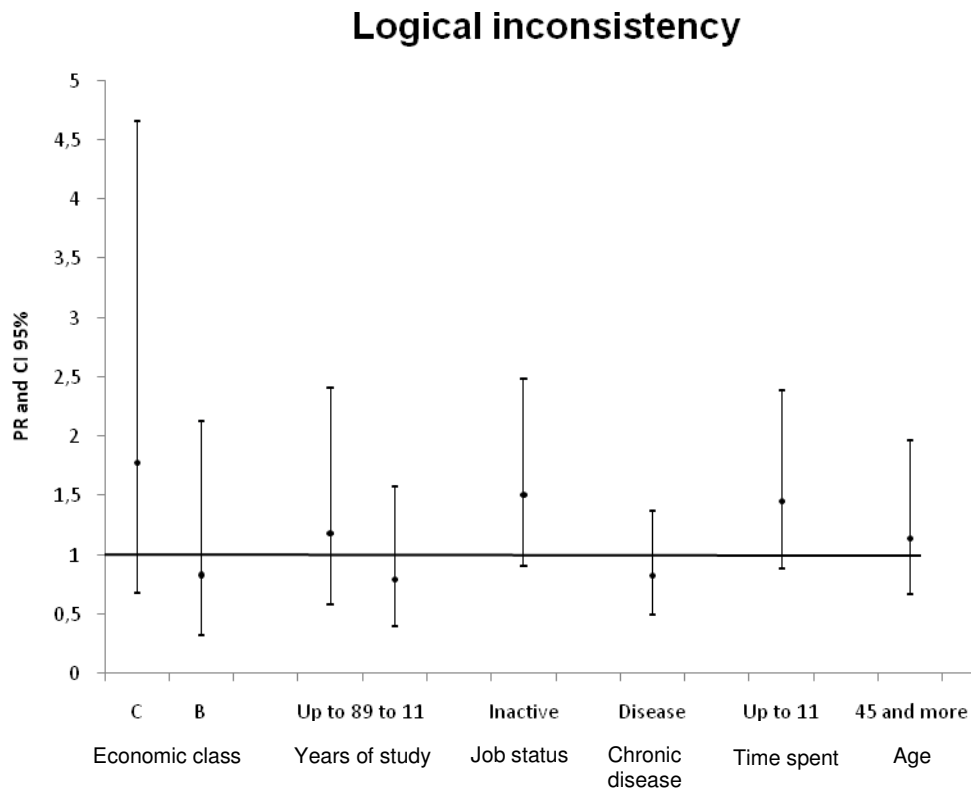


Figure 4 - Representation of regression model 3, showing Prevalence Ratios and CI 95% among logical inconsistency and consistent responses according to the following independent variables (from left to right): economic class, years of study, job status, presence or not of disease, time spent in the standard gamble (in minutes) and age.

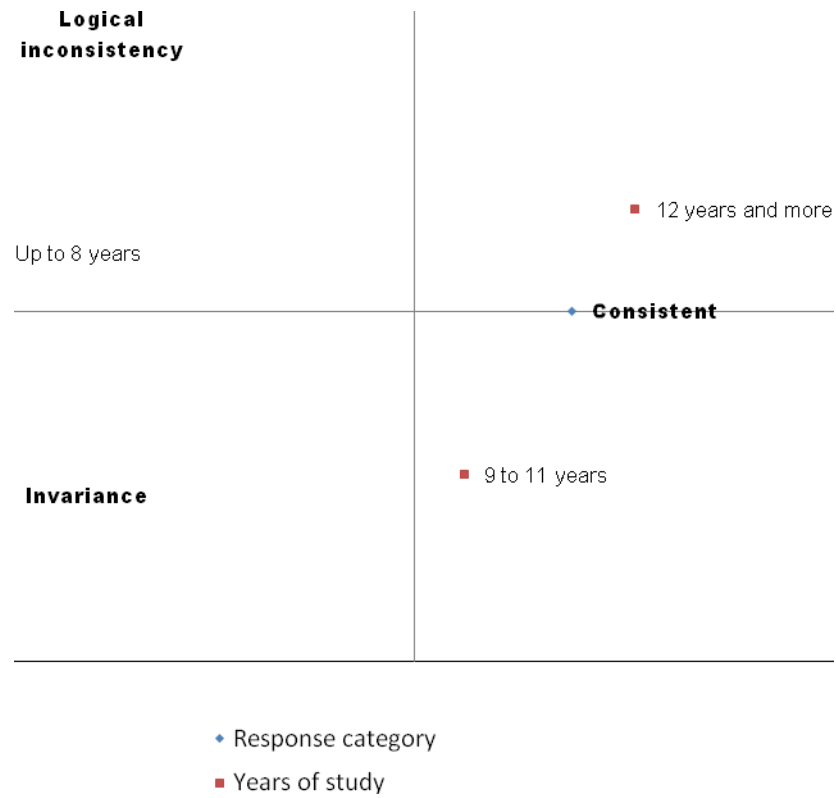


Figure 5 - Correspondence analysis between the response type (invariance, logical inconsistency and consistency) and the education variable, in years of study.

ARTIGO 3

**ESTIMATING THE SF-6D VALUE SET FOR A
POPULATION BASED SAMPLE OF BRAZILIANS**

ESTIMATING THE SF-6D VALUE SET FOR A POPULATION BASED SAMPLE OF BRAZILIANS

Luciane N. Cruz ¹, Suzi A. Camey ², Juliana F. Hoffmann¹, Donna Rowen³, John E. Brazier³, Marcelo P. Fleck⁴, Carisi A. Polanczyk^{1, 5}

Artigo a ser submetido para o periódico Value in Health

Graduate Studies Program in Epidemiology ¹, Statistics Department², Federal University of Rio Grande do Sul, Porto Alegre, Brazil; School of Health and Related research, University of Sheffield, UK³; Graduate Program in Psychiatry⁴, Federal University of Rio Grande do Sul, Porto Alegre, Cardiology Division of Hospital de Clinicas de Porto Alegre⁵, Brazil.

Corresponding author:

Luciane Nascimento Cruz

Rua Ramiro Barcelos, 2350/second floor

Porto Alegre – RS – Zip Code 90035-003

Phone: 55 51 3356-7880

Fax: 55 51 3342-7959

E-mail: Incruz@terra.com.br

ABSTRACT

Objectives: The SF-6D is a preference-based measure of health developed to estimate utility values from the SF-36. The aim of this study was to estimate a weighting system for the SF-6D health states representing the preferences of Brazilian general population.

Methods: A sample of 248 health states defined by the SF-6D was valued by a sample of a southern Brazilian population using the standard gamble (SG) method. The SG responses were used to estimate regression models at the individual and mean levels to predict preference values for all SF-6D health states. The models were compared with those described in the UK study.

Results: Five hundred twenty eight participants were interviewed, but 146 (28%) were excluded due to inconsistent SG responses. Data from 382 subjects were used to estimate the models, rendering 2224 health states valuations. Compared to UK data, Brazilian health state values were lower, leading to a lower constant term in the models. The best fit model for the Brazilian data was a random effect model using only the main effects variables, different from the British SF-6D model, highlighting the importance of adopting a country-specific algorithm in predicting SF-6D health states values. The Brazilian model produced significant coefficients and presented a mean absolute difference between observed and predicted values of 0.07. Inconsistent coefficients were merged to produce the final recommended model.

Conclusion: The results provide the first population-based value set for health states in Brazil, making it possible to generate QALYs for cost-utility studies using regional data. Besides, utility scores based on Brazilian preferences values can be derived from existing SF-36 data set.

RESUMO

Objetivos: O SF-6D é uma medida baseada em preferências desenvolvida para calcular índices de utilidade a partir do SF-36. O objetivo deste estudo foi estimar um sistema de valores para os estados de saúde gerados pelo SF-6D que representam as preferências da população geral em uma cidade brasileira.

Métodos: Uma amostra de 247 estados de saúde definidos pelo SF-6D foi avaliado por uma amostra da população geral da cidade de Porto Alegre, utilizando o método do *standard gamble* (SG). As respostas ao SG foram utilizadas para estimar modelos de regressão capazes de extrapolar os valores para todos os possíveis estados de saúde gerados pelo SF-6D. Os modelos foram comparados com aqueles descritos no estudo original do Reino Unido.

Resultados: Quinhentos e vinte oito participantes foram entrevistados, mas 146 (28%) foram excluídos devido a respostas inconsistentes ao *standard gamble*. Os dados de 382 indivíduos foram utilizados para estimar os modelos, resultando em 2224 avaliações dos estados de saúde. Em comparação com dados do Reino Unido, os valores atribuídos aos estados de saúde pela população do estudo foram mais baixos, levando a um baixo valor do intercepto nos modelos. O modelo mais bem ajustado aos dados coletados no presente estudo foi um modelo de efeito aleatório, utilizando apenas as variáveis de efeitos principais, diferente do modelo britânico, destacando a importância de adotar um algoritmo específico para cada país. O modelo produziu coeficientes significativos e apresentou uma diferença média absoluta entre os valores observados e estimados de 0,07. Coeficientes inconsistente foram agregados para produzir o modelo final recomendado.

Conclusão: Os resultados do estudo fornecem a primeira base de dados populacional para preferências por estados de saúde no Brasil, possibilitando o cálculo de QALYs para estudos de custo-utilidade utilizando dados regionais. Além disso, índices de utilidade poderão ser derivados de bancos de dados do SF-36.

INTRODUCTION

Increasingly, decision makers, providers, patients and the public require that expenditures on health be justified according to expected outcomes. In this context, the decision-making process in health and healthcare policy has never been more important to reduce inefficiencies, eliminate ineffective medical procedures, increase competitiveness, improve quality, change reimbursement formulas and ration services (1). Cost-effectiveness analysis is increasingly used to aid in the process of decision making for resource allocation in health. An important tool in this analysis is the Quality Adjusted Life Year (QALY), an index that combines quantity and quality of life (2). International guidelines for studies of cost-effectiveness (3) and institutions of health technology assessment as the National Institute of Health and Clinical Excellence (NICE) (4) in the UK have recommended that the QALYs are the reference outcome for economic evaluation.

Health-related quality of life (HRQOL) measures suitable for calculating QALYs are those that incorporate preferences into their scoring system, the so-called preference-based measures. Examples of widely used generic instruments are the EQ-5D(5), Health Utilities Index(6) and SF-6D (7). The latter was developed for obtaining a preference-based index from the SF-36 for use in economic evaluation.

The growing number of studies conducting health states valuation surveys around the world have contributed to cross-cultural comparisons. Studies valuing the EQ-5D (8-11) and SF-6D (12-14) in diverse cultures have shown that health state preferences are different from the preferences derived in the country where the measure was originally created. This information reinforces the importance of each culture generating its own set of preference weights for health states.

The aim of this study was to estimate preference weights for SF-6D health states that represent the preferences of a sample of southern Brazilian general population. This article presents results of the valuation survey and the modeling of the SF-6D health states, comparing them with those from the original UK study.

METHODS

This study follows the same protocol as the original UK SF-6D valuation study (7). First, the SF-36 was reduced in size and complexity in order to generate health states to be valued by respondents. Second, a preference valuation survey in the general population was undertaken. Finally, the results of the survey modeled to predict the values for all possible health states generated by the reduced version of SF-36 (SF-6D).

Deriving the SF-6D health state classification

With the aim of deriving the descriptive system of the Brazilian SF-6D, we used the same derivation of SF-36 items used in the original study of the United Kingdom (7), employing the version of the SF-36 translated and validated in Brazil by Ciconelli et al. (15). The result was a 6 dimension descriptive system, each with multiple levels: 1) physical functioning: 6 levels, 2) role limitations: 4 levels, 3) social functioning: 5 levels, 4) pain: 6 levels, 5) mental health: 5 levels; 6) vitality: 5 levels. This descriptive system is identical to the UK classification system except the language.

Selection of health states

A health state was defined by selecting one level of each dimension. Considering the number of dimensions and levels, the SF-6D generates 18,000 health states ($6 * 4 * 5 * 6 * 5 * 5$). We used the same 247 health states valued in the original UK study to enable comparisons. States were classified as mild, moderate or severe. Each individual rated 5 intermediate states plus the worst health state generated by SF-6D (the “pits” state). Care was taken to ensure that each respondent was asked to value health states covering a wide range of severity according to the SF-6D classification system (7).

Sampling

The sample consisted of individuals selected from the general population of Porto Alegre, a capital city in southern Brazil. The sample size was calculated such that each of the 247 health states was valued at least 10 times. Since each individual would evaluate 5 intermediate states, the sample size estimation was 494 respondents.

A sample of the urban general population of Porto Alegre aged 20 to 64 years was selected. A two-stage cluster random selection design was used. Primary units were represented by census sectors of the municipality according to division made by IBGE (Brazilian Geography & Statistics Institute)(16). In a second stage, households were systematically selected in each census sector. The number of households to be visited was calculated considering the average number of adults per household and the population's proportion in each one of the strata the study aimed to reach, that is, men and women in the age ranges of 20-29, 30-44, 45-64 years. The interviewers visited 7 domiciles in each of the 108 selected sectors, inviting all residents who met the following inclusion criteria to take part in the study: aged 20 to 64 and literate and with no physical or mental incapacity that would prevent reading and understanding required in the valuation tasks.

Study protocol

A team of interviewers made up of graduation students from health-related courses received intensive training to apply the standard gamble technique and the remainder of the interview protocol. The study was first introduced by means of a letter that described the purpose of the study, the type of task that the interviewer would develop and the time spent in the interview, which was estimated at about 40-60 minutes. Interviews were later conducted in the respondents' own home.

At the start of the interview respondents completed the SF-6D, in order to familiarize respondents with the idea of describing health in terms of the SF-6D. At the next stage of the interview, respondents were asked to rank, according to their preferences, a group of eight cards containing the 5 intermediate states along with the best state defined by the SF-6D, the worst state (the "pits") and immediate death.

The main task of the interview was the valuation of 5 intermediate health states plus the worst state using the standard gamble (SG) technique. Respondents were asked to choose between the certain prospect of living in an intermediate state defined by the SF-6D and the uncertain prospect of two possible outcomes, the best state defined by the SF-6D or the worst state (the “pits”). The chances of the best state occurring was varied systematically during the exercise, until the respondent was indifferent between the certain and uncertain prospect. As a visual aid to apply SG, a graphic material (“props”) developed by a team from the Mac Master University (Furlong, 1990) and made in Brazil with the permission of the authors was used. This chance board shows the chances of the two uncertain outcomes occurring, both numerically and in the form of a pie chart. The “ping-pong” version of SG was used as this makes the interview more didactic, leading the interviewer through a set of questions depending on the respondents answer to the previous question, standardizing the procedure and minimizing the risk of variation between interviewers.

Following the valuation of the 5 intermediate states, respondents valued the worst SF-6D state (pits) using a modified version of the standard gamble. The SG choice depended on how the respondent had ordered the worst state, if better or worse than death, in the ranking exercise: 1) if better than death, the respondent must choose between the certain prospect of the worst state and the uncertain prospect of full health or immediate death; 2) if worse than death, the choice was between the certain prospect of death and the uncertain prospect of full health or the worst state. The choice process and props are identical to those used to value intermediate states. The SG value produced from this SG task is measure on the full-health 1-0 scale used to produce QALYs. All values derived for the intermediate states are then adjusted onto the full health-dead 1-0 scale using the pits value for each individual. Taking P to represent the value given to the worst state for individual i and SG to represent the values of an intermediate state j , the formula used to generate adjusted values $SGADJ$ for all intermediate health state valuations is:

$$SGADJ_{ij} = SG_{ij} + (1-SG_{ij}) * P_i$$

These values were used in the estimation of the models.

At the end of the standard gamble, respondents were asked how difficult they found the tasks using a Likert scale: “very easy”, “easy”, “somewhat difficult”, “difficult” and “very difficult”.

In the next stage of the interview, respondents completed the Short-Form 36 (SF-36) (17) and the World Health Organization Quality of Life Instrument-Brief (WHOQOL-)(18), both translated and validated for use in Portuguese (15;19). The study described in this paper is part of a broader project that, in addition to valuation of the SF-6D health states, will produce normalization of scores of the SF-36 and WHOQOL-brief instruments in the general population of a southern Brazilian city. For this reason, these instruments were included in the protocol description, but the data will not be presented here.

The last part of the protocol consisted of a socio-demographic questionnaire applied by the interviewer. All respondents signed the Informed Consent Form approved by the University’s Institutional Review Board.

Variables

The socio-demographic variables obtained were: gender, age, marital status (married and not married), level of education (in years of study), job status (employed, unemployed, informal job, housewife, student, retired and other), and economic class.

Economic class was evaluated through an index called the Brazil Criterion that classified the population into 7 classes according the purchasing power and level of education of the head of the family (16). The division and its equivalence in terms of monthly family income in dollars are, approximately: Class A1: 3, 800; Class A2: 2,300; Class B1: 1,400; Class B2: 800; Class C: 460; Class D: 212; & Class E: 103.

Descriptive statistics of all health state values were generated and compared to UK values.

Modeling

Regression analysis was used to estimate preference weights for each level and dimension of the classification system to enable preference weights to be

estimated for all states rather than simply those included in the valuation study. This analysis followed the same protocol as the UK study, using linear models to estimate the relationship between the SF-6D classification system and standard gamble adjusted values obtained in the valuation survey. Models were estimated at both the individual and aggregate (mean values for the 248 health states) levels.

Several models were constructed to predict the health states values, but the general individual level model can be defined by the formula:

$$y_{ij} = g(\beta \cdot x_{ij} + \theta r_{ij}) + \varepsilon_{ij} \quad (1)$$

where $i = 1, 2, \dots, m$ represents the valued health state and $j = 1, 2, \dots, n$ represents respondents. The dependent variable, y_{ij} , is the standard gamble adjusted score for the health state i valued by respondent j (SGADJ), x is a vector of dummy independent variables ($x_{\delta\lambda}$) for each level λ of dimension δ do SF-6D. For example, x_{31} means that the variable refers to the dimension 3 at level 1. For any health state, $x_{\delta\lambda}$ is defined as:

$$x_{\delta\lambda} = 1, \text{ if, for this state, dimension } \delta \text{ is at level } \lambda \ (\lambda \geq 2)$$

$$x_{\delta\lambda} = 0, \text{ if, for this state, dimension } \delta \text{ is not at level } \lambda$$

Overall there are 25 terms, with the level $\lambda = 1$ acting as a baseline for each dimension. The coefficient of the dummy variables represents the main effect of a move from level 1 to the other levels in each dimension. Assuming a simple linear model, the intercept represents the state of perfect health (111111). The final value of each health state can be estimated by summing the coefficients of the levels of each dimension present in a given state.

The r term is a vector to account for the interactions between the levels of different dimensions, g is a function specifying the appropriate functional form; ε_{ij} is an error term whose autocorrelation structure and distribution depends on the assumptions of each model estimated.

The first model was an OLS (ordinary least square) regression model, considering g as a linear function. This specification assumes zero mean error and

constant variance error structure with independent error terms. Each individual health state valuation was considered an independent observation, regardless of whether it was valued by the same individual.

The second model was a random effects model, which takes account of variation both within and between respondents. For this model, the error term of formula (1) is subdivided so that:

$$\varepsilon_{ij} = u_j + e_{ij} \quad (2)$$

where u_j is the respondent variation and e_{ij} is an error term for the i th valuation of the j th individual. A random variation is assumed for both terms. Estimation was through restricted maximum likelihood (RMLE).

Additional strategies were undertaken to deal with the possible effects of interaction between the levels of different dimensions of the SF-6D. Dummy variables were created to take account of any additional effect on health state value when one or more dimensions are at the extreme levels. Least severe was defined as level 1 or 2 on each dimension. Most severe was defined as levels 4-6 for physical functioning, levels 3 and 4 for role limitation, levels 4 and 5 for social functioning, mental health and vitality and levels 5 and 6 for pain in line with UK models. This procedure resulted in the dummy variables "LEAST" and "MOST". These variables had a value equal to 1 if any dimension in the health state was at least (LEAST) or most (MOST) severe levels and equal to 0 otherwise.

The models were evaluated considering the following criteria: a) inconsistencies in the coefficients estimates: the coefficients of dummy variables representing each level of SF-6D are expected to be negative and increasing in absolute size as the level of severity increases (amongst coefficients with statistical significance); b) the mean absolute error (MAE), and the proportion of predictions outside 0.05 (% AE > 0.05) and 0.10 (% AE > 0.10) ranges on either side of the observed value. Predictions were further tested in terms of bias (t test), regarding to the normality of residuals (Jarque-Bera - JB) and the presence of autocorrelation in the prediction errors (statistics Ljung-Box-LB). Analysis was performed using SPSS version 16.0 (IBM company, Chicago) (20), R 2.9.1 (Vienna) (21) and STATA 9.0 (Stata Corp., Texas) (22).

RESULTS

Study population

A total of 889 households were visited from July 2007 to March 2008. Interviews were conducted in 354 (40%) of them, because for the others it was impossible to contact dwellers after 3 consecutive visits or people refused to welcome the study team. Out of 846 eligible individuals identified, 537 accepted to take part, obtaining a response rate of 64%, similar to that of 65% found in the population of the UK study. The number of eligible individuals includes all the households in which the research team was able to contact, even households where people refused to participate, because we could collect information regarding the number of people aged from 20 to 64 living in the contacted addresses.

Out of 537 respondents, 9 (1.6%) were excluded because they did not complete the ranking exercise or the standard gamble, leaving a sample of 528 respondents. The same exclusion criteria used in the UK were then applied to select the respondents whose values would be used in the modeling analysis. Respondents were excluded who: a) had not valued the worst state, since without this value it was not be possible to adjust the values of all other intermediate states; b) had not valued two or more health states; c) had given the same value to the 5 intermediate states valued.

All respondents valued the worst health state, but for 58 respondents (11%), the response was considered invalid. Invalid values represented one of two situations: 1) the individual chose to remain in the worst health state for the rest of his (her) life rather than accept the gamble that offered 100% chance of returning to full health or; 2) SG preference for the worst health state was inconsistent with the ordering of death and worst state in the ranking task. For example, in the ranking task, the respondent considered the worst state as better than death, but in the SG task preferred chose to accept the gamble offering 100% chance of dying. Since this response violates the expected logic, those giving it were excluded from the final sample.

There were 88 respondents (17%) excluded because their values for the 5 health states did not vary. The lack of variation is likely to indicate that the respondent did not understand the task.

The comparison between excluded and included subjects according to socio-demographic characteristics is presented in Table 1. The 146 excluded subjects were older, of a lower socioeconomic class and less educated.

The responses of the 382 respondents included in the final sample, 68 were missing, resulting in a completion rate of 97%. This final sample results in 2224 observed standard gamble valuations across 248 health states.

Health state values

Descriptive statistics for 40 of the 248 health states are shown in Table 2, comparing the values observed in the Brazilian population with those of the original UK study. Overall, the Brazilian mean health state values were lower and range from 0.16 to 0.88 with large standard deviations. Similar to UK data, median health state values generally exceed mean values, reflecting a negative skewness of the data.

Figure 1 presents a histogram for the 2224 individual adjusted health state valuations. As found in the UK study, negative values (states considered worse than death) were comparatively rare (Brazil 6.6% and UK 6.9%) and the proportion of valuations assigned to the maximum value (1.0) was small (Brazil 0.4% VS UK 0.5%). Regarding to the worst health state, 63% of respondents valued it as better than death, while in the UK the rate was 73%. However, 12% of the Brazilian observations range between 0.9 and 1.0 in comparison to more than 23% observed in the UK sample.

Modeling

Several models were estimated following the strategy proposed by Brazier et al.(7), but only the best models in terms of predictive ability are described in this article. Therefore, the results are presented for the random effects models with and without the intercept restricted to unity (Table 3). In these models the variables created to take account of interaction effects, "LEAST" and "MOST", did not improve the predictive ability of the models and for this reason are not presented here.

In the random effects model without the intercept forced to unity (model 1 in Table 3) 13 of 25 coefficients were significant with four inconsistencies, where the

estimated effect decreases from level 2 to level 3 for the physical functioning and mental health dimensions and coefficients for SF2 and vitality dimension did not show the expected negative sign, except the most severe level, which in turn was not significant. In terms of predictive ability, the proportion of prediction errors under 0.1 and 0.05 was 84% and 53% respectively. The predictions are unbiased ($p > 0.05$), but prediction errors are not normally distributed (JB test). Moreover, there is autocorrelation in the prediction errors (LB statistics), as can be seen in Figure 2, which shows the observed and predicted values for model 1. There is a tendency to over-predict at low health state values and under-predict at high health state values.

According to the axioms of expected utility theory underlying the standard gamble technique, the best health state generated by the SF-6D (111111) is equal to 1 and death is equal to zero. For state (111111) to hold any other value would not conform to the conventional utility scale. The best way to ensure that the best state has the value 1 is to restrict the intercept to unity (7) .

Table 3 shows the random effects model (model 3) with the constant forced to unity. There was an increase in the number of significant coefficients compared to the previous model 1, but a higher number of inconsistencies. All coefficients had the expected negative sign. Regarding the number of prediction errors model 3 performed worse than model 1, with a higher proportion with absolute errors greater than 0.05 and 0.1. As in the model 1, the predictions were unbiased and the residuals were not normally distributed. An important advantage of model 3 was the absence of autocorrelation in the prediction errors (LB test not significant). Figure 3 shows the curves of observed and predicted values for the 248 health states valued using model 3.

Consistent models

Additional strategies were used to estimate models without inconsistencies, following an approach used in the UK SF-6D (23). Consistent models were constructed from models 1 and 3 by removing variables that were not significant at $p < 0.05$ and aggregating levels of each dimension if inconsistencies occurred; that is if the coefficients on each level did not represent an additional decrease to health state value. The results are shown in Table 3. Model 2 is a consistent version of

model 1, where levels 2 and 3 of physical functioning dimension have been merged because the coefficients are not significant. From level 4 to level 5 of the same dimension there was an inconsistency, since the coefficient of PF4 was higher than PF5 in absolute value. So these two levels were merged as well. Coefficient for SF2 have been merged with the level 1 reference point because it was insignificant and inconsistent (positive sign), as well as the coefficients of levels 2, 3 and 4 of vitality dimension. In model 3, all coefficients were significant but inconsistent coefficients were merged (PF2 and PF3; RL2 and RL3; PAIN2 and PAIN3; MH2 and MH3 and VIT2, VIT3 and VIT4). This procedure generated models 2 and 4. Both are similar in terms of predictive ability compared to their original models. Model 4 have a higher mean absolute error (MAE) and higher proportion of predictions outside 0.05 and 0.1 ranges related to the model 3, but has the advantage of having the intercept fixed at 1 according to the conventional utility scale. Furthermore, the model does not suffer from autocorrelation in the errors of prediction. For these reasons, this model appears to be the most appropriate. Figure 4 shows the observed and predicted values for model 4.

Table 4 shows a comparison between the Brazilian consistent model (model 4) and the UK consistent model used to derive SF-6D (7). Statistically, there are not comparable models, since the UK model was estimated by Ordinary Least Squares (OLS) using data at a mean level, including the interaction term (“MOST”). The OLS models with interaction terms applied to the study data performed worse than all other models estimated in terms of predictive ability. MAE increased to 0.102 and the percentage of predictions errors outside 0.05 was above 70% and 42% outside 0.01, with problems of autocorrelation in the errors. The interaction term (MOST) was not significant.

The coefficients in the models can be interpreted as representing any decrement in utility associated with health deteriorating from full health. In the recommended model 4, the value of full health is equal to one, since the intercept was forced to the unity. Utility values for all possible SF-6D health states can be obtained by using the coefficients estimates by subtracting from 1 the utility decrement associated with each level in a given health state. For example, taking the state 245633, the estimated value would be: $1 - 0.079 - 0.081 - 0.120 - 0.155 - 0.087 - 0.021 = 0.478$.

DISCUSSION

The results of this study provide the first population-based value set for health states in Brazil, making it possible to generate QALYs for cost-utility studies using regional data. This is a relatively new research area in Latin America, with only one study valuing EQ-5D health states in the population of Argentina, recently published (11). The Brazilian SF-6D preference weights estimated here offer a method for deriving preference values from existing SF-36 data.

Since the overall aim was to construct a model to predict values for all possible health states generated by the SF-6D, the main criterion used to choose the most appropriate model was the predictive ability in terms of mean absolute errors between observed and predicted health state values and percentages of prediction errors above 0.05 and 0.01. On this basis, the random effects models were the best among all models tested. On theoretical grounds it is argued that the constant term in the models should be set equal to one to conform with the conventional utility scale when 1 represents perfect health and 0 represents death. This would suggest model 3 as the preferred model.

One concern is the existence of inconsistencies between coefficients, meaning that the utility value does not reduce as health becomes more severe. This finding can be related to the difficulty in attempting to value a comparatively large classification system describing 18,000 health states. Due to sample size issues, some values are not very stable. This result is similar to that found by the authors of the original study and other researchers in different countries (7;12;14). It is likely that increasing the number of health states to be evaluated, which in turn require a larger number of individuals, may overcome some of the inconsistencies and no significant coefficients for some levels in some dimensions(23).

Moreover, the inconsistencies may be expected due to the classification system of the SF-6D. For example, there is an inconsistency between level 4 ("Your health limits you a lot in moderate activities") and level 5 ("Your health limits you a little in bath and dressing ") of the physical functioning dimension. Perhaps some respondents have not considered the level 5 worse than level 4, or the expression "a lot" on the level 4 has caused a greater impact to the respondent. Between levels 2 and 3 of the mental health dimension another inconsistency was observed, which

may be caused by the difficulty of distinguishing between "you feel tense or downhearted and low a little of the time" (level 2) or "... some of the time "(level 3). The same inconsistency in the vitality dimension (between levels 2 and 3) was found in the UK study and the authors' interpretation was that the description of this dimension may have caused some confusion for respondents, since it is worded in the positive rather than negative ("You have a lot of energy most of the time", "You have a lot of energy some of the time) (7).

As the task of carrying out a population survey is highly costly, in terms of time and resources, an alternative for dealing with inconsistent coefficients was proposed by Brazier et al (23). Following this strategy, it was possible to construct consistent models by merging inconsistent coefficients. Consistent models were estimated from the models 1 and 3, containing only the variables that contribute significantly to the final health state value. Therefore, the recommended model for calculating utility values for the SF-6D health states in the Brazilian population is model 4.

The differences between Brazilian and UK models reinforce the importance of using a country-specific algorithm for calculations of utility values in the national context. The best fitting model in the Brazilian data was a random effects consistent model that takes into account variations in two levels, inter-and intra-individual, while for the UK data the best model was an ordinary least square consistent model using data at a mean level. The coefficients of the Brazilian model were larger than those found in the UK, except for the vitality dimension, leading to a greater decrement in utility values from full health. While the bodily pain dimension appears to be the most important dimension in determining the health state value for both cultures, the physical functioning and social functioning dimensions seems to have greater importance for the Brazilians. These differences between the weights of different dimensions among diverse cultures was also found in other countries regarding SF-6D (12-14), and similar instruments such as EQ-5D (8-10;24), emphasizing the need for more studies analysing the association between socio-cultural variables and preference measurement.

One concern regarding our results is that the number of individuals included in the analysis was lower than the estimated 494 participants. The decision to exclude the 146 inconsistent respondents was a trade-off between the decision to keep the estimated sample size and to construct a model containing only values considered

consistent. Some findings can be influenced by sample size issues. For example, the results showed that mild states (21111) presented a mean value (0.69), equal to the value of an intermediate state 423343 (0.69), suggesting that respondents did not seem to differentiate mild states from better ones. It could be speculated that respondents in this survey were willing to take much higher risks for avoiding mild health states, potentially reflecting the risk attitudes of the study sample. However, this can be a result by chance due to the lower number of health states valuations. On the other hand, we do not believe that the same problem occurred with modeling. Using data at individual level to build the models increase the degrees of freedom, because the observational unit is the health state valuations. So, the estimates of the model coefficients are based in the number of valuations for each level of each domain. Probably, this can be an explanation for the fact that our models using data at an individual level performed much better than models using data at a mean level in terms of predictive ability.

Another important aspect to be highlighted is that the sample used is not representative of all regions of Brazil. Due to the great cultural diversity, this research should be replicated in the other regions so that the national data set become available.

While we recognize some caveats of the study, it is important to emphasize the difficulty of conducting a population survey in our country. Due to high rates of urban violence, in many census sectors people use to live in buildings with security systems that greatly hinder access to residents. For these reasons, it was necessary to adopt the strategy of replacement of losses and refusals, visiting a number of households larger than planned in order to obtain the required number of interviews.

The availability of a regional algorithm for calculation of utility index represents an opportunity to undertake local health economics research in Brazil. To date, studies of cost-utility in Brazil were conducted using secondary data derived from other countries, generally developed ones. This study estimated preference weights using a random sample of the general population, making suitable for incorporation in the decision-making process for resource allocation and public health policies in Brazil.

Acknowledgements: Dr. Luciane Cruz received a graduate research scholarship from CAPES, Brazil. Prof. Polanczyk and Prof. Fleck received a research scholarships from CNPq/Brazil

This study was funded by CNPQ/Brazil (Edital MCT-CNPq / MS-SCTIE-DECIT – N^o 36/2005).

References

- (1) Patrick DL, Erickson P. Applications of Health Status Assessment to Health Policy. In: Spilker B, ed., *Quality of Life and Pharmacoeconomics in Clinical Trials* (Second ed.). Philadelphia: Lippincott-Raven, 1996. p. 717-27.
- (2) Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health* 2009 Mar;12 Suppl 1:S27-S30.
- (3) Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996 Oct 16;276(15):1253-8.
- (4) National Institute for Health and Clinical Excellence. National Institute for Health and Clinical Excellence (NICE) 2010 March 7 Available from: URL: www.nice.org.uk.
- (5) EuroQol - a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy* 1990 Dec;16(3):199-208.
- (6) Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001 Jul;33(5):375-84.
- (7) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002 Mar;21(2):271-92.
- (8) Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 2005 Jun;6(2):124-30.
- (9) Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002 Jun;11(4):341-53.
- (10) Johnson JA, Luo N, Shaw JW, et al. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med Care* 2005 Mar;43(3):221-8.
- (11) Augustovski FA, Irazola VE, Velazquez AP, et al. Argentine valuation of the EQ-5D health states. *Value Health* 2009 Jun;12(4):587-96.
- (12) Brazier JE, Fukuhara S, Roberts J, et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol* 2009 Dec;62(12):1323-31.

- (13) 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 Mar;11(2):295-303.
- (14) Ferreira LN, Ferreira PL, Pereira LN, et al. A Portuguese Value Set for the SF-6D. *Value in Health*. In press 2010.
- (15) Ciconelli RM, Ferraz MB, Santos W, et al. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras Reumatol* 1999;39(3):143-50.
- (16) Censo Populacional 2000. Instituto Brasileiro de Geografia e Estatística 2010 March 6 Available from: URL: www.ibge.gov.br.
- (17) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 Jun;30(6):473-83.
- (18) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 1998 Jun;46(12):1569-85.
- (19) Fleck MP, Louzada S, Xavier M, et al. [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref]. *Rev Saude Publica* 2000 Apr;34(2):178-83.
- (20) Statistical Package for Social Sciences. IBM Company 2010 March 6 Available from: URL: <http://www.spss.com/software/>.
- (21) R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2010 March 6(2009) Available from: URL: www.R-project.org.
- (22) Data analysis and statistical software. StataCorp LP 2010 March 6 Available from: URL: www.stata.com.
- (23) Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004 Sep;42(9):851-9.
- (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 Jan;21(1):7-16.

Table 1 - Socio-demographic characteristics of the included and excluded subjects

Variables	Included (n=382) %	Excluded (n=146) %	<i>p</i>
Age (mean \pmSD)	41 \pm 12	44 \pm 13	0.02
Gender			
Male	41	37	0.36
Female	59	63	
Marital status			
Married	63	60	0.42
Not married	37	40	
Economic class			
A1	1	2	0.007
A2	15	7	
B1	19	14	
B2	25	21	
C	34	48	
D	6	8	
Years of study			
Up to 4 years of study	4	8	< 0.001
5 to 8 years	16	31	
9 to 11 years	35	32	
12 years or more	45	29	
Job status			
Employed	54	44	0.28
Informal job	8	9	
Unemployed	7	5	
Housewife	14	21	
Student	4	6	
Retired	7	10	
Other	5	5	

Tabela 2 - Descriptive statistics for 40 health state valuations comparing Brazil and UK (7)

State	Brazil					UK				
	Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD
111112	0.42	1.00	0.81	0.94	0.21					
214535	0.50	0.94	0.80	0.83	0.16	0.00	0.99	0.63	0.78	0.37
423343	0.19	1.00	0.69	0.81	0.29	0.00	1.00	0.44	0.38	0.31
211111	0.30	1.00	0.69	0.73	0.28	0.19	1.00	0.78	0.90	0.27
213323	0.06	0.92	0.67	0.80	0.28	0.12	0.98	0.74	0.79	0.25
111215	0.10	1.00	0.66	0.96	0.43	0.53	1.00	0.90	0.97	0.14
412152	0.25	0.95	0.66	0.77	0.29	0.10	0.93	0.50	0.59	0.29
131542	0.29	0.96	0.64	0.65	0.23	-0.66	0.96	0.42	0.45	0.41
224223	0.44	0.95	0.64	0.58	0.20	0.53	1.00	0.80	0.85	0.17
443215	0.21	0.98	0.63	0.70	0.27	-0.06	1.00	0.67	0.81	0.35
241531	0.19	0.96	0.61	0.62	0.29	0.28	0.99	0.75	0.88	0.24
134322	-0.23	0.97	0.59	0.67	0.37	0.10	1.00	0.57	0.59	0.27
112221	0.00	0.96	0.58	0.85	0.43	0.51	0.98	0.82	0.89	0.17
345122	0.15	0.88	0.58	0.65	0.27	0.29	1.00	0.64	0.67	0.25
122233	-0.40	0.94	0.57	0.78	0.45	0.14	1.00	0.83	0.91	0.23
221432	0.05	0.96	0.57	0.52	0.30	0.53	0.98	0.81	0.84	0.15
312552	0.19	0.99	0.55	0.51	0.37	0.10	0.95	0.53	0.64	0.35
342353	0.05	1.00	0.55	0.67	0.39	0.29	0.98	0.66	0.79	0.23
344145	-0.28	0.98	0.50	0.66	0.41	-0.57	0.98	0.51	0.63	0.48
112521	-0.48	0.94	0.49	0.57	0.35	0.19	0.94	0.71	0.73	0.21
321455	0.05	0.95	0.48	0.45	0.29	0.10	0.99	0.60	0.65	0.33
423433	0.00	0.91	0.47	0.53	0.37	-0.15	1.00	0.58	0.60	0.36
523634	0.19	0.64	0.46	0.53	0.19	0.05	0.99	0.61	0.57	0.33
541432	0.19	0.75	0.46	0.47	0.20	0.10	1.00	0.69	0.75	0.29
432623	0.06	0.98	0.44	0.35	0.29	0.07	1.00	0.55	0.56	0.30
141653	0.15	0.72	0.43	0.44	0.25	0.00	0.91	0.39	0.36	0.34
431623	-0.14	0.80	0.43	0.66	0.39	-0.88	0.99	0.45	0.67	0.47
341123	0.15	0.80	0.41	0.32	0.27	0.10	0.99	0.76	0.92	0.31
323333	-0.88	0.99	0.40	0.57	0.57	0.05	0.98	0.68	0.76	0.32
323644	0.09	0.75	0.40	0.34	0.28	0.10	0.99	0.40	0.29	0.31
315515	0.02	0.75	0.39	0.38	0.25	0.19	0.97	0.56	0.55	0.25
241635	0.18	0.67	0.38	0.40	0.18	-0.09	0.99	0.54	0.57	0.37
124314	0.15	0.70	0.36	0.30	0.24	0.06	0.99	0.70	0.94	0.35
532124	0.05	0.85	0.36	0.37	0.28	0.29	1.00	0.79	0.84	0.21
534644	0.05	1.00	0.36	0.23	0.33	-0.28	0.98	0.35	0.32	0.32
334254	0.10	0.93	0.31	0.19	0.30	-0.66	0.98	0.58	0.80	0.46
535645	0.05	0.70	0.30	0.19	0.26	-0.56	0.76	0.10	0.10	0.39
434654	-0.28	0.81	0.28	0.14	0.34	-0.85	1.00	0.38	0.55	0.61
432255	-0.38	0.97	0.24	0.15	0.43	0.00	1.00	0.48	0.48	0.42
645655	-0.98	0.98	0.14	0.05	0.38					

SD=Standard deviation; Min=Minimum; Max= Maximum

Table 3 - Main effects models and consistent models^a

	Constant forced to unity					
	RE ^b	Consistent model		RE	Consistent model	
	(Model 1)	(Model 2)		(Model 3)	(Model 4)	
c	0.671		0.705	1.000		1.000
PF2	-0.020			-0.080		
PF3	-0.030			-0.068	PF23	-0.075
PF4	-0.063	PF23	-0.024	-0.105		
PF5	-0.048	PF45	-0.053	-0.070	PF45	-0.084
PF6	-0.142	PF6	-0.145	-0.183	PF6	-0.187
RL2	-0.018			-0.067		
RL3	-0.032	RL23	-0.025	-0.063	RL23	-0.064
RL4	-0.049	RL4	-0.046	-0.084	RL4	-0.079
SF2	0.010			0.056	SF2	-0.058
SF3	-0.016	SF3	-0.024	-0.068	SF3	-0.064
SF4	-0.057	SF4	-0.061	-0.110	SF4	-0.108
SF5	-0.074	SF5	-0.078	-0.118	SF5	-0.117
PAIN2	-0.017			-0.100		
PAIN3	-0.014			-0.064	PAIN23	-0.084
PAIN4	-0.023	PAIN234	-0.023	-0.082	PAIN4	-0.086
PAIN5	-0.045	PAIN5	-0.053	-0.112	PAIN5	-0.112
PAIN6	-0.109	PAIN6	-0.111	-0.152	PAIN6	-0.154
MH2	-0.046			-0.095		
MH3	-0.034	MH23	-0.043	-0.079	MH23	-0.084
MH4	-0.054	MH4	-0.053	-0.096	MH4	-0.091
MH5	-0.100	MH5	-0.102	-0.142	MH5	-0.135
VIT2	0.048			-0.031		
VIT3	0.013			-0.037		
VIT4	0.030			-0.027	VIT234	-0.032
VIT5	-0.028	VIT5	0.056	-0.077	VIT5	-0.051
n	2224		2224	2224		2224
Inconsistencies	4			5		
MAE	0.060		0.060	0.070		0.072
%AE > 0.05	47		50	51		54
%AE > 0.10	16		16	25		25
t (mean=0)	0.181		1.291	0.980		1.055
JBPREP	187.61		183.53	35.254		25.46
LB	12.80		11.28	1.095		0.625

a Estimates shown in bold are significant at $p < 0.05$

b RE= Random Effects

Table 4 - Consistent models for Brazil and UK

	Brazil		UK	
	Consistent model ¹		Consistent model ²	
	(Model 4)			
c		1.000		1.000
PF2				
PF3	PF23	-0.075	PF23	-0.035
PF4			PF4	-0.044
PF5	PF45	-0.084	PF5	-0.056
PF6	PF6	-0.187	PF6	-0.117
RL2				
RL3	RL23	-0.064		
RL4	RL4	-0.079	RL234	-0.053
SF2	SF2	-0.058	SF2	-0.057
SF3	SF3	-0.064	SF3	-0.059
SF4	SF4	-0.108	SF4	-0.072
SF5	SF5	-0.117	SF5	-0.087
PAIN2				
PAIN3	PAIN23	-0.084	PAIN23	-0.042
PAIN4	PAIN4	-0.086	PAIN4	-0.065
PAIN5	PAIN5	-0.112	PAIN5	-0.102
PAIN6	PAIN6	-0.154	PAIN6	-0.171
MH2				
MH3	MH23	-0.084	MH23	-0.042
MH4	MH4	-0.091	MH4	-0.100
MH5	MH5	-0.135	MH5	-0.118
VIT2				
VIT3				
VIT4	VIT234	-0.032	VIT 234	-0.071
VIT5	VIT5	-0.051	VIT5	-0.092
			MOST	-0.061
n		2224		249
Inconsistencies				
MAE		0.072		0.074
%AE > 0.05		54		47
%AE > 0.10		25		23
t (mean=0)		1.055		-1.317
JBPRED		25.46		1.169
LB		0.625		163.87

¹ Model estimated from a random effects model with data at the individual level

² Model estimated from an ordinary least square model at a mean level including interaction term.

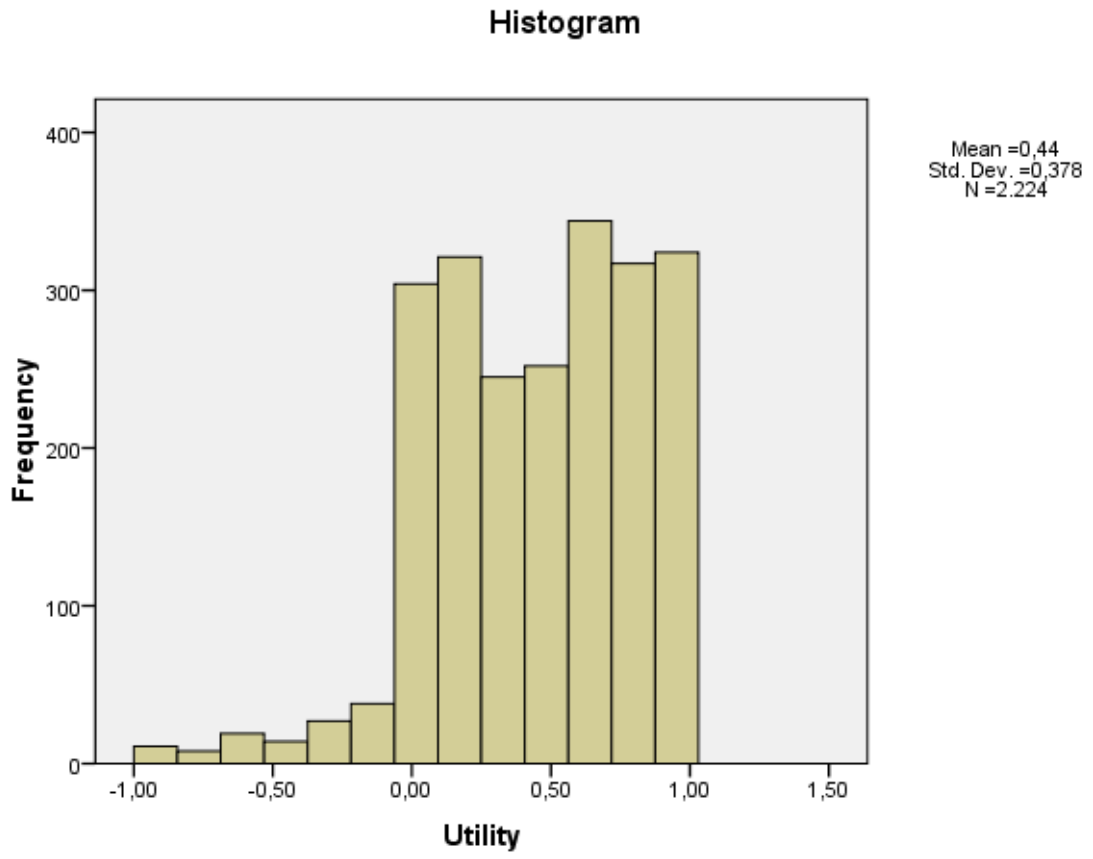


Figure 1 - Histogram for adjusted health state valuation

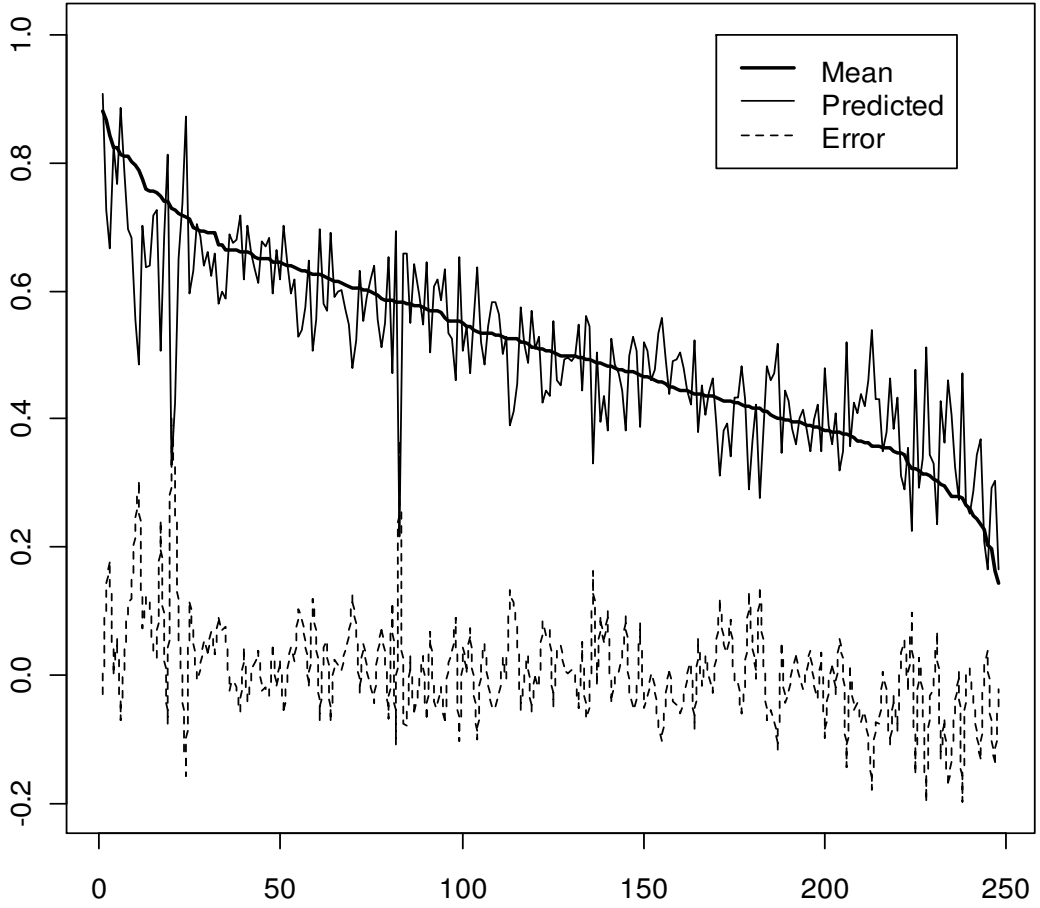


Figure 2 - Observed and predicted health state valuation for the random effects model (1).

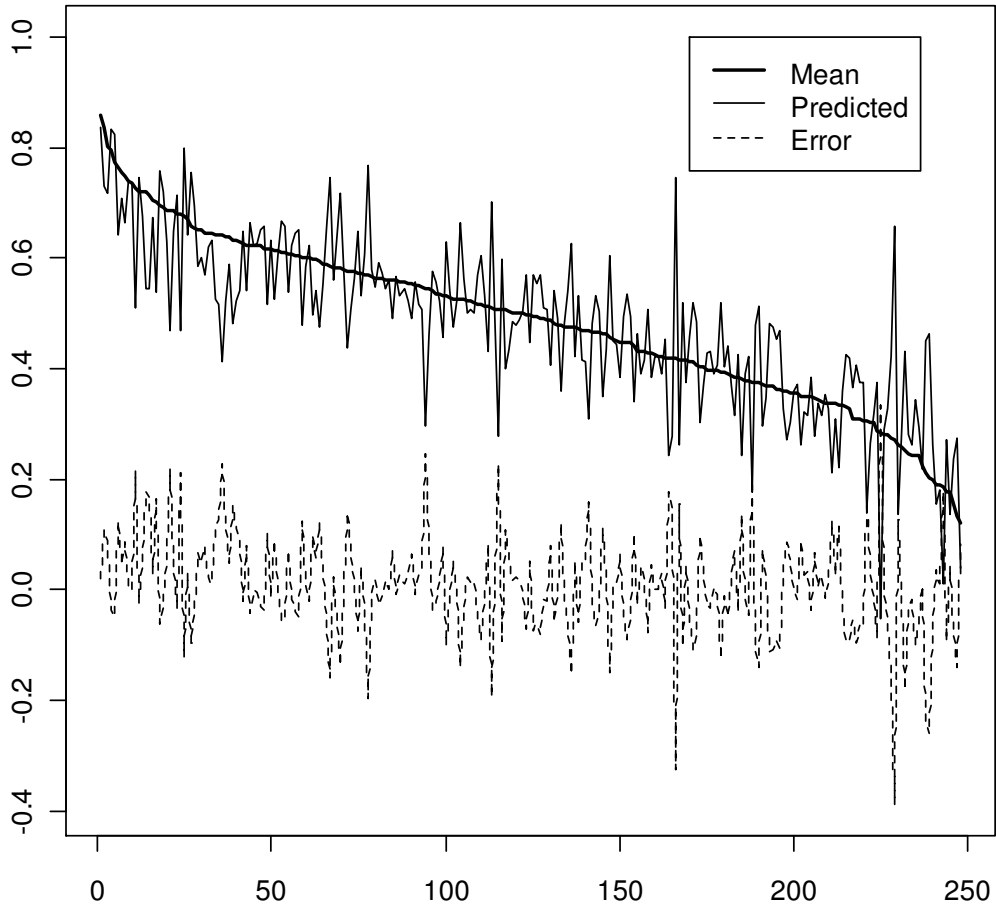


Figure 3 - Observed and predicted health state valuation for the random effects model with constant forced to unity (3).

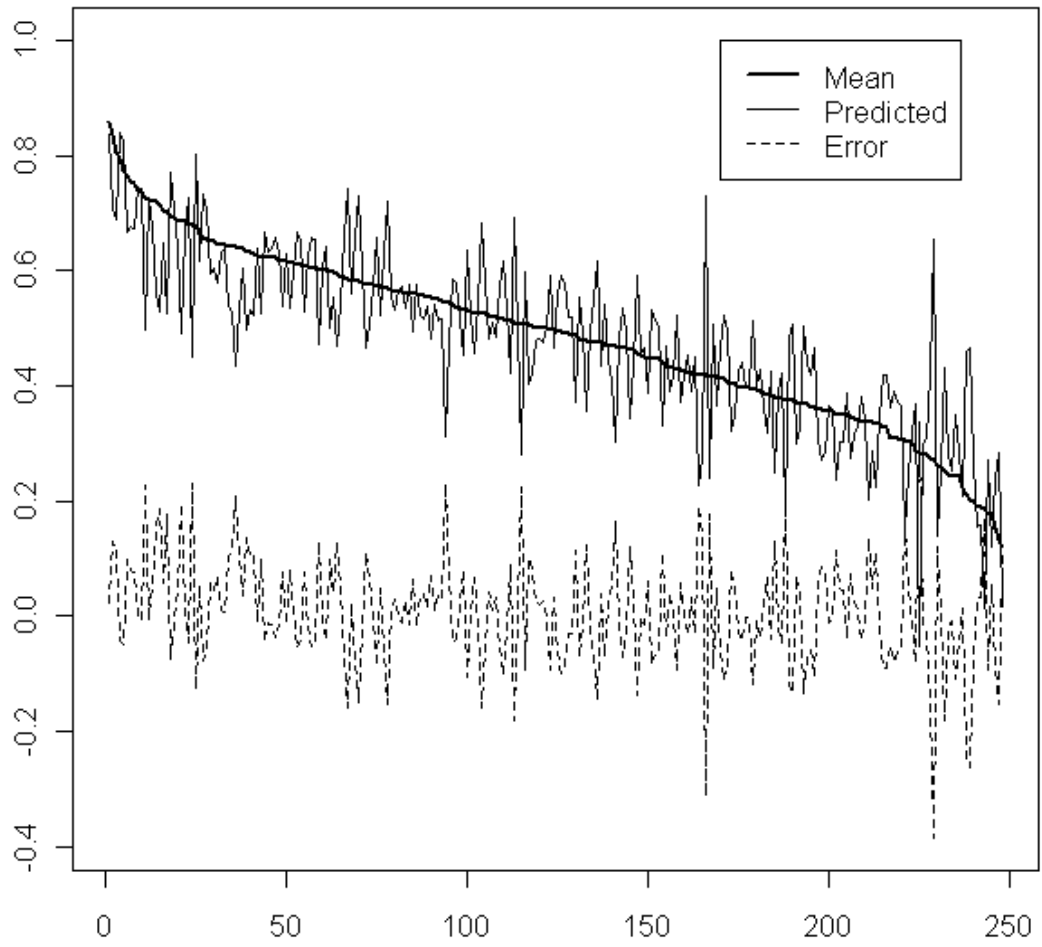


Figure 4 - Observed and predicted health state valuation for the consistent random effects model with constant forced to unity (4).

ARTIGO 4

**HEALTH-RELATED QUALITY OF LIFE IN BRAZIL:
NORMATIVE DATA FOR THE SF-36 IN A SOUTHERN
GENERAL POPULATION SAMPLE**

ARTIGO 4

**HEALTH-RELATED QUALITY OF LIFE IN BRAZIL:
NORMATIVE DATA FOR THE SF-36 IN A SOUTHERN
GENERAL POPULATION SAMPLE**

Luciane N. Cruz ¹, Marcelo P. Fleck², Suzi A. Comey³, Juliana F. Hoffmann¹,
Carisi A. Polanczyk^{1,4}

Artigo a ser submetido ao periódico Journal of Clinical Epidemiology

Graduate Studies Program in Epidemiology ¹, Graduate Studies Program in
Psychiatry², Statistics Department³, Federal University of Rio Grande do Sul, Porto
Alegre, Brazil; Cardiology Division of Hospital de Clinicas de Porto Alegre⁴ , Brazil.

Corresponding author:

Luciane Nascimento Cruz

Rua Ramiro Barcelos, 2350/2nd floor

Porto Alegre – RS – Zip Code 90035-003

Phone: 55 51 3356-7880

Fax: 55 51 3342-7959

E-mail: Incruz@terra.com.br

ABSTRACT

Objective: To provide normative scores of SF-36 in a general population sample in Brazil and to describe differences in mean scores according to sociodemographic characteristics. **Study Design and Setting:** SF-36 was applied to a randomly selected sample of the general population of Porto Alegre. Participants were literate people aged 18 to 64 years. The questionnaires were self-administered in the presence of an interviewer in the respondent's home. **Results:** The response rate was 68% and 755 subjects were included (38% male, 62% female). 99% of respondents fulfilled 100% of the SF-36 questions. Lower health status was reported to females, to the 30-44 year age cohort, to subjects in a lower economic class, less educated and reporting a chronic medical condition. Means and percentiles of scores of the SF-36 are reported as normative data for the general population. **Conclusion:** SF-36 was an acceptable and practical instrument for measuring health related quality of life in a sample of general population in Porto Alegre. The results of this study can be useful to researchers using the SF-36 questionnaire in other cohorts to compare scores with normative data. SF-36 can be a valuable tool for studies on health inequalities due to the ability to discriminate different population sub cohorts.

RESUMO

Objetivo: Disponibilizar escores normativos do SF-36 em uma amostra da população geral no Brasil e descrever as diferenças de escores médios de acordo com características sociodemográficas. **Método:** SF-36 foi aplicado a uma amostra aleatória da população geral de Porto Alegre. Os participantes eram pessoas alfabetizadas com idade entre 18 a 64 anos. Os questionários foram auto-administrados, na presença de um entrevistador no domicílio do entrevistado. **Resultados:** A taxa de resposta foi de 68% e 755 indivíduos foram incluídos (38% homens, 62% do sexo feminino). 99% dos entrevistados preencheram 100% das questões do SF-36. Pior estado de saúde foi registrado para mulheres, em indivíduos na faixa etária de 30 a 44 anos, de uma classe econômica mais baixa, com menor escolaridade e com auto-relato de uma condição médica crônica. Médias e percentis dos escores do SF-36 são apresentados como dados normativos para a população geral. **Conclusão:** O SF-36 foi um instrumento aceitável e prático para medir a qualidade de vida relacionada à saúde em uma amostra da população geral em Porto Alegre. Os resultados deste estudo podem ser úteis para pesquisadores utilizando o SF-36 em outros grupos para comparar os escores com dados normativos. O SF-36 pode ser uma valiosa ferramenta para detectar grupos vulneráveis em estudos epidemiológicos, devido à capacidade de discriminar diferentes subgrupos da população.

INTRODUCTION

The importance of quality-of-life assessment (QOL) has been expressively growing in the past 50 years. Some factors that have contributed to the increased use include the accumulation of evidence that it is a valid and reliable measure, the publication of clinical studies demonstrating that these measures are responsive to clinical changes, and the development of shorter instruments easier to use and understand (1).

One of the most used health-related quality of life (HRQOL) instrument worldwide is the Medical Outcomes Study Short-Form 36 (SF-36) (2). The SF-36 was created from the necessity of having a standardized instrument that would address general health concepts not specific for any medical condition, and that were understandable, easy to use and psychometrically appropriate. The conceptual basis for the development of SF-36 were the concepts of functional status and well-being described in accepted definitions of "health" (2). Thus, the concept of quality of life considered in the elaboration of this instrument was the health-related quality of life, emphasizing the specific impact that prevention and treatment of a disease has on the "value of being alive".

The health concepts assessed by the SF-36 are: physical functioning, social functioning, role functioning, general health and mental health perceptions, pain and vitality. As a generic instrument, it is useful for comparing general and specific populations, comparing the relative impact of diseases, differentiating the benefits produced by different treatments and screening individual patients (3).

SF-36 has been translated into several languages and adapted to several cultures. The International Quality of Life Assessment (IQOLA) is the project of a group of researchers from Europe and the United States where the guidelines for the translation and cultural adaptation process of SF-36 were delineated, which consists of 3 stages: 1. Translation; 2. Psychometric evaluation of the items; 3. Empirical validation and norming of scores (4). Normative data enable the interpretation of scores of the instrument for an individual or the average of a group, since there is no "gold standard" against which to compare the results obtained with this instrument.

Population norms are available for many developed countries (4), but there are a limited number of studies reporting these data in developing ones (5;6). In Latin America, translations and validations of the SF-36 are available for a few countries (7;8), but this is the first study, of our knowledge, to report population normative data in this context. This seems to be an important research question, since Brazil is the only Latin American country that speaks Portuguese and adds in its territory several ethnic groups and cultures, thus requiring regional normative data for comparison of health-related quality of life scores.

METHODOLOGY

Sampling

The sample consisted of individuals selected from the general population of Porto Alegre, a capital city in the South of Brazil. This is a city with 1,436,123 inhabitants, being the capital of one of the most developed states of the country, with 97% of the population living in urban areas, per capita GDP of approximately US\$ 13,000.00 and a literacy rate of 96.7% (9).

The estimated sample size was 800 individuals, according to the minimum sample size recommended by IQOLA project (4).

A two-stage cluster random selection design was used. In a first stage of sampling, a random sample of 108 census sectors of the city was obtained, divided by the Brazilian Institute of Geography and Statistics (IBGE). To calculate the number of households to be visited, the average number of adults per household was considered, and the population's proportion in each one of the strata the study aimed to reach, that is, men and women in the age ranges of 20-29, 30-44, 45-64 years. In each sector, 7 households to be visited were systematically selected and all residents were invited to participate in the study if they met the following inclusion criteria: age ranging from 20 to 64 years; be literate; not having any physical or mental limitation that could prevent the reading and understanding of the instruments. If the residents were not found in the first visit, another two visits in different days and times, including non-business hours, were carried out. A cover letter containing the team identification and purposes of the study, time to be spent on the interview and phone

numbers was provided to residents present in the first visit, or deposited in the mailbox of the selected households.

Instruments

SF-36 is a generic instrument whose conceptual basis is “health-related quality of life”. This construct is represented by 36 questions divided into eight domains: physical functioning, role physical, pain, general health, vitality, role social, role emotional and mental health. Items are scored by a Likert scale. All items of SF-36 are used to score the eight domains, except for item 2, which refers to a self-report of health transition. Each item contributes to only one domain. After recalibrating two items and reverse the score of nine items, the responses to items are summed. The highest scores represent better health status. One score for missing values is computed if items of one scale are not responded. Scores range from 0 to 100, 0 indicating the less favorable health status and 100 the most favorable one. SF-36 is a questionnaire that can be administered by: self-administration, administration by computer, personally or by phone calls by a trained interviewer and is adequate for individuals above 14 years of age. It can be administered in 5 to 10 minutes with high degree of acceptability and quality of data (3).

The SF-36 employed in this study was previously translated into Portuguese and validated in Brazil by Ciconelli and cols. (7). The study was performed with a population of patients with rheumatoid arthritis, using the protocol elaborated in compliance with some steps proposed by the IQOLA coordinators (10).

A standardized questionnaire was used to obtain socio-economic and demographic data and contained the following variables: gender, age, race, marital status, practice of any religion, employment status and economic class, number of medical consultations and admissions in the last year, smoking and alcohol use. The presence of chronic diseases was assessed by a list of diseases with dichotomous response (yes/no): hypertension, diabetes, ischemic cardiopathy (infarct/angina), heart failure, arthrosis/arthritis, cerebrovascular accident, chronic bronchitis/emphysema, asthma, kidney disease, cancer, HIV/AIDS, back pain, depression and anxiety, and one open question codified as “others”.

The economic class was assessed by an index called Brazil Criterion (Critério Brasil) which divides the population into classes according to their purchasing power and schooling of the family head (11). The classification and its equivalence concerning approximate mean family income in American dollars would be: Class A1: Mean family income of US\$ 3,800; Class A2: US\$ 2,300; Class B1: US\$ 1,400; Class B2: US\$ 800; Class C: US\$ 460; Class D: US\$212, and Class E: US\$ 103.

Statistical analysis

Continuous data are expressed as means \pm standard deviation and categorical in percentage. Comparisons of QOL mean scores among groups according to socio-demographic characteristics were performed by ANOVA, Brown-Fosythe or t-test. ANOVA was used when there was homogeneity of variances and Brown-Fosythe when there were not both of them for comparisons between two or more groups. Levene's test was used to perform tests of homogeneity of variances.

For all tests a significance level was established at <0.05 . Data were analyzed using SPSS for Windows, version 13.0 (IBM Company, Chicago) and Microsoft Office Excel 2003.

RESULTS

From July/2007 to July/2009, 1057 households were visited, being possible to perform the interviews in 514 (49%), because for the others it was impossible to contact dwellers after 3 consecutive visits or people refused to welcome the study team. From 1119 eligible individuals, identified and contacted, 758 participated in the project, achieving a response rate of 68%. The number of eligible individuals includes all the households in which the research team was able to contact, even households where people refused to participate, because we could collect information regarding the number of people aged from 20 to 64 living at the addresses contacted.

It was necessary to exclude 3 individuals from the sample, 2 (0.3%) for error in the age record and 1 (0.1%) for not responding more than 50% of the SF-36 items, totaling 755 participants with data available for analysis.

In order to assess the representativeness of the sample, its composition was compared to the population of the city of Porto Alegre, as shown in Table 1. The distribution in terms of gender and age group of the sample was similar to that of the general population, except for the subgroup represented by males from 30 to 44 years that had a smaller percentage of individuals. In relation to economic class, there was a subrepresentativity of lower classes, D and E, probably due to some criteria used by the study protocol such as exclusion of illiterates, which are usually included in these strata of the population. Additionally, 8 (6.7%) census sectors had to be excluded from the sample because they are places with high rates of urban violence that could jeopardize the safety of the team members. Such exclusions, because they involve neighborhoods possibly inhabited by people of lower purchasing power, may also justify the low level of D class and the lack of E class.

The socio-demographic characteristics of the sample are described in Table 2. The mean age of the sample was 41 ± 13 years and 62% of the participants were females. Respondents were mostly married, white, practiced a religion, and with formal employment. The mean of study years was 11.3 ± 5.1 , and 37% of the sample with 12 years or above of study.

Forty nine percent of participants reported having some chronic medical condition, the most common being hypertension (13.5%), arthritis (8.3%), asthma (7.5%), and diabetes mellitus (4.6%). Depression was reported by 14% of the sample and anxiety by 21%.

Seven hundred forty-eight participants (99%) responded to all questions of the SF-36. The items with higher number of missing values (1%) were PF4 ("Climbing several flights of stairs" – physical functioning domain) and RP3 ("Accomplished less than would like" – role physical domain). The mean of time spent to answer to SF-36 was 10 ± 5.2 minutes.

Descriptive statistics for the 8 domains of SF-36 are available in Table 3. As expected in data obtained from a general population sample, most of the respondents scored in the favorable health scores, a finding observed through the high median seen in all domains, and the negative asymmetry, indicating the trend of scores to the upper range of the scale. This finding is also evidenced by the high percentage of ceiling effects, that is, respondents scoring at the highest score, especially in the role physical and role emotional domains. A high index of ceiling

effect was also seen in the social functioning domain. On the other hand, in a sample of the general population, the percentage of participants with scores at the lower scale should be minimal, as seen in the present study. From the 8 domains, 1.5% or less of floor effect was observed in 6. There was exception in the role physical and role emotional domains, which had higher percentages of individuals with minimum scores, 11 and 16%, respectively.

Normative data for the 8 domains of SF-36 according to gender, age group, economic class, educational level and presence of disease are available in Table 4. Results are presented by mean and standard deviation of the scores.

Women had the worst health status, with statistically significant difference ($p < 0.001$ to 5 domains) in all domains. Major differences occurred in pain and vitality domains and the minor ones in general health and mental health domains. In terms of age, statistically significant differences between the mean scores were found only in domains related to physical health (physical functioning, pain, role physical and general health), with decreasing values as increasing age.

The mean scores also varied according to economic class and educational level. Values decreased in a proportional manner to education, with statistical significance in most of the areas, except for the pain, social functioning and role emotional domains. A worst health status was also seen in individuals of lower economic classes, with statistically significant differences in scores in all domains, except pain and role emotional.

As a measure of health status, SF-36 was able to differentiate the group of individuals that reported having some chronic health condition from the group that considered itself healthy, with worst health status for the first group. The differences in scores had statistic significance ($p < 0.001$) in all the 8 domains. The major discrepancies occurred between means of individuals who reported having depression and those who not reported, mainly in the role emotional and mental health, as expected.

Table 5 shows the scores in each area of the SF-36 in percentiles 5, 10, 25, 50, 75, 90, 95 for the total sample and for each subgroup according to gender and age group. The description in percentiles is to make the use of scores here available more practical for future comparisons.

Figure 1 shows the mean scores for the 8 domains of SF-36 obtained in this research compared to the normative scores of other 4 countries with different cultures. Brazil has lower scores than developed countries and Turkey in nearly all domains, except for vitality, where the score was higher in relation to all the other. Comparing to Croatia, a developing country, the studied Brazilian population presented higher mean scores.

DISCUSSION

The results of our research provide regional normative data for SF-36 to be used by researchers in comparisons of cohorts of individuals in different clinical situations. In the absence of “gold standards” for health measures, normative scores can be very useful in interpreting scale scores for an individual respondent or the average score for a group in comparison to the distribution scores for individuals from the general population (4).

This study sought to meet the requirements recommended by the guidelines for standardization of scores of the SF-36. The number of participants was near to the 800 individuals as suggested by IQOLA and the study complied with other criteria such as response rate over two thirds, demographic information including age, sex, employment status, education, marital status and a checklist of self-reported chronic conditions (4). In spite of slightly smaller sample size, important to detect differences in mean scores between groups, the results of the present study regarding the ability to differentiate individuals according to demographic variables and presence of disease were similar to those found in other countries which used a larger number of individuals in the sample (5;6;12-18).

The quality of data of the study was high, considering as criterion the percentage of missing values for items and domains of the SF-36, which was below 2% (19). This percentage was lower than that found in Medical Outcomes Study (MOS), a study that used the original version of the SF-36, which ranged from 1.1 to 5.9% (19). The authors of SF-36 emphasize that the scores cannot be estimated with the same confidence level if there is a large number of missing data (20). Additionally, the non-response index also reflects the understanding and acceptance of the questionnaire by the participants (19). In this sample of the general population

of Porto Alegre, SF-36 seems to have been well accepted, and it was of quick application, with mean of completion time of 10 minutes.

The low number of missing values in this research might have been secondary to the administration mode of SF-36, which was self-applicable, but performed in the presence of the interviewer, who checked if all questions were responded. The few missing items that occurred were probably due to a refusal of the respondent to complete that item.

The distribution of SF-36 scores in the total of sample is comparable to that found in the application of the original instrument in the general population of the United States (3), with most respondents having higher scores. The areas with highest percentages of floor and ceiling effects were the same, role physical and role emotional. These two domains are considered the “coarsest” of the eight scales, enumerating only five or four levels of health each. One of the ways to improve this limitation of these areas would be to replace dichotomous responses by responses with more categories that measure finer gradations in role disability aside from the mere presence or absence of limitation (19). The fact that the highest level of functioning is merely defined by the absence of physical or emotional limitations causes the ceiling effect in the above areas to always be a limitation in the SF-36 application in samples of non-diseased individuals. Younger individuals also had more domains with the highest score in relation to older individuals, confirming a possible reduction in sensitivity at the upper limits of the scale in people with less functional limitations. Data for the SF-36 in populations of patients with chronic diseases had lower prevalence of ceiling effect (21).

The observed differences in mean scores of the SF-36 among different population strata emphasize the need to use the standards described for each subgroup for comparison. The main discrepancies were related to gender, women presenting a worse health status in all domains of the SF-36. This finding seems to be independent from culture and socioeconomic status, since it was unanimous in normalization studies conducted in different countries of Western Europe (14-18), Canada (12), New Zealand (22) and Mexico (23). For the other socio-demographic variables, older individuals reported a worse health status only in domains related to physical health, while respondents with less education and lower socioeconomic class had the lowest scores in almost all areas. These findings were also seen in

other studies performed on developed countries (15), and developing ones (5;24), demonstrating the advantage of using the SF-36 in populational studies to identify groups of vulnerable individuals. The description of health-related quality of life in different areas also allows identifying which aspects of the life of the individual might be more affected. The graphic showing the curves of scores in different countries showed that the population of some countries reported better health in the physical areas, but worse in areas such as vitality, for example. Additionally, since this is a generic instrument, it enables cross-cultural comparisons.

SF-36 was able to clearly differentiate between the subjects with self-reported diseases and the group declared healthy, suggesting good construct validity of this instrument developed to measure health status. The group that reported to have some chronic condition had worst health status in all areas. Individuals who identified themselves with depression and anxiety, the only two psychiatric conditions included in the protocol, had mean scores significantly lower in the 8 domains, with major differences in role emotional and mental health.

One important aspect to be highlighted is that the sample used is not representative of all regions of Brazil. Due to the great cultural diversity, this research should be replicated in the other regions so that the national normative tables become available.

Another caveat of this study is the fact that the sample presented a subrepresentativity of the general population in relation to the lower socioeconomic classes. Some criteria followed by the protocol such as to exclude illiterate and preserve the team to work on areas of greatest risk for urban violence may have led to a smaller representation of D and E classes. Since the quality of life scores decreased progressively in lower socioeconomic classes, one can infer that the classes D and E would score even lower. Then, using the results presented in this paper to compare quality of life between groups, researchers should be aware that for individuals pertaining to classes D and E, values are probably overestimated.

While we recognize some caveats of the study, it is important to emphasize the difficulty of conducting a population survey in our country. Due to high rates of urban violence in our city, many people use to live in buildings with security systems that greatly hinder access to residents. For these reasons, it was necessary to adopt the strategy of replacement of losses and refusals, visiting a number of households

larger than planned in order to obtain the required number of interviews.

To conclude, SF-36 seems to be an acceptable and easily applicable instrument to the general population, and its performance proved to be similar to that found in other general population samples around the world. It is a useful tool to measure the health status in cross-sectional epidemiological studies, but it has limitations in some scales to detect positive changes in health status in longitudinal studies of populations without chronic diseases. The normative values available in this study can be used as reference for comparison of scores obtained from different cohorts of patients.

Acknowledgements: We thank the Research Incentive and Event Fund of Hospital de Clínicas de Porto Alegre for the financial aid in translating this article. Dr. Luciane Cruz received graduate research scholarship from CAPES, Brazil. Prof. Polanczyk and Prof. Fleck received a research scholarship from CNPq/Brazil

This study was funded by CNPQ/Brazil (Edital MCT-CNPq / MS-SCTIE-DECIT – Nº 36/2005).

References

- (1) Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995 Jan 4;273(1):59-65.
- (2) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 Jun;30(6):473-83.
- (3) Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976)* 2000 Dec 15;25(24):3130-9.
- (4) Gandek B, Ware JE, Jr. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998 Nov;51(11):953-9.
- (5) Demiral Y, Ergor G, Unal B, Semin S, Akvardar Y, Kivircik B, et al. Normative data and discriminative properties of short form 36 (SF-36) in Turkish urban population. *BMC Public Health* 2006;6:247.
- (6) Maslic SD, Vuletic G. Psychometric evaluation and establishing norms of Croatian SF-36 health survey: framework for subjective health research. *Croat Med J* 2006 Feb;47(1):95-102.
- (7) Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Revista Brasileira de Reumatologia* 1999;39(3):143-50.
- (8) Augustovski FA, Lewin G, Elorrio EG, Rubinstein A. The Argentine-Spanish SF-36 Health Survey was successfully validated for local outcome research. *J Clin Epidemiol* 2008 Dec;61(12):1279-84.
- (9) IBGE Cidades. Instituto Brasileiro de Geografia e Estatística (IBGE). Available from: URL: www.ibge.gov.br. Last update: 12.02.10
- (10) Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998 Nov;51(11):913-23.
- (11) Critério Econômico Brasil 2003. Associação Brasileira de Empresas de Pesquisa. Available from: URL: www.abep.org/novo/default.aspx. Last update 12.02.10
- (12) Hopman WM, Towheed T, Anastassiades T, Tenenhouse A, Poliquin S, Berger C, et al. Canadian normative data for the SF-36 health survey. Canadian Multicentre Osteoporosis Study Research Group. *CMAJ* 2000 Aug 8;163(3):265-71.

- (13) Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993 May 29;306(6890):1437-40.
- (14) Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998 Nov;51(11):1055-68.
- (15) Sullivan M, Karlsson J. The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: results from normative population. *J Clin Epidemiol* 1998 Nov;51(11):1105-13.
- (16) Alonso J, Regidor E, Barrio G, Prieto L, Rodriguez C, de la Fuente L. [Population reference values of the Spanish version of the Health Questionnaire SF-36]. *Med Clin (Barc)* 1998 Oct 10;111(11):410-6.
- (17) Bowling A, Bond M, Jenkinson C, Lamping DL. Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *J Public Health Med* 1999 Sep;21(3):255-70.
- (18) Pappa E, Kontodimopoulos N, Niakas D. Validating and norming of the Greek SF-36 Health Survey. *Qual Life Res* 2005 Jun;14(5):1433-8.
- (19) McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994 Jan;32(1):40-66.
- (20) Gandek B, Ware JE, Jr., Aaronson NK, Alonso J, Apolone G, Bjorner J, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998 Nov;51(11):1149-58.
- (21) Cruz LN, Camey SA, Fleck MP, Polanczyk CA. World Health Organization quality of life instrument-brief and Short Form-36 in patients with coronary artery disease: do they measure similar quality of life concepts? *Psychol Health Med* 2009 Oct;14(5):619-28.
- (22) Scott KM, Tobias MI, Sarfati D, Haslett SJ. SF-36 health survey reliability, validity and norms for New Zealand. *Aust N Z J Public Health* 1999 Aug;23(4):401-6.
- (23) Duran-Arenas L, Gallegos-Carrillo K, Salinas-Escudero G, Martinez-Salgado H. [Towards a Mexican normative standard for measurement of the short format 36 health-related quality of life instrument]. *Salud Publica Mex* 2004 Jul;46(4):306-15.

- (24) Wang R, Wu C, Zhao Y, Yan X, Ma X, Wu M, et al. Health related quality of life measured by SF-36: a population-based study in Shanghai, China. *BMC Public Health* 2008;8:292.

Table 1 - Socio-demographic comparison between total sample and the general population of Porto Alegre

	Total sample (n=755) N (%)	General Population * (%)
Men		
20-29 years	102 (14)	15
30-44 years	74 (10)	16
45-64 years	112 (15)	15
Women		
20-29 years	97 (13)	15
30-44 years	138 (18)	19
45-64 years	228 (30)	20
Economic class		
A1	9 (1)	1
A2	77 (10)	5
B1	116 (15)	7
B2	189 (25)	17
C	307 (41)	38
D	52 (7)	28
E		5

Table 2 - Socio-demographic characteristics of the study population

Variables	N (Total=755)	%
Age (mean - SD)	41.3 (12.9)	
Gender		
Male	290	38
Female	465	62
Race*		
White	607	81
Black	74	10
Brown	54	7
Others	19	2.5
Religion *		
Yes	404	53.5
No	349	46.2
Marital status		
Married	447	59
Single	195	26
Widowed	25	3
Divorced	88	12
Years of study*		
Up to 4 years of study	46	6
5 to 8 years	168	22
9 to 11 years	259	34
12 years or more	278	37
Job status (%)		
Employed	381	50
Informal job	62	8
Unemployed	50	7
Housewife	117	16
Student	41	5
Retired	69	9
Other	35	5

*Total amount is not 755 due to missing values

Table 3 - Descriptive statistics for the eight domains of SF-36 for the total sample

	Physical Functioning	Role Physical	Bodily Pain	General Health	Vitality	Social Functioning	Role Emotional	Mental Health
Mean	82.45	74.73	67.53	71.10	66.85	78.30	70.02	73.82
SD	20.51	35.33	23.75	19.52	20.39	23.88	38.12	19.79
Median	90	100	72	72	70	87	100	80
Skeweness	-1.57	-1.09	-0.50	-0.66	-0.89	-1.04	-0.83	-1.09
Floor (%)	0.3	11.0	1.5	0.3	0.5	0.9	15.9	0.1
Ceiling (%)	24.8	58	17.7	4.4	2.8	38.8	55.0	5.4

Table 4 - Mean scores of SF-36 domains according to sociodemographic characteristics and presence of self-reported diseases

	Physical Functioning		Role Physical		Bodily Pain		General Health	
	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P
Gender								
Male	87.1 (17.2)	<0.001	79.5 (31.4)	0.002	73.8 (21.6)	<0.001	73.7 (18.4)	0.003
Female	79.4 (21.8)		71.7 (37.3)		63.6 (24.2)		69.4 (20.0)	
Age (years)								
20-29	91.4 (11.8)	<0.001	79.7 (30.5)	0.019	74.2 (20.9)	<0.001	75.0 (17.4)	0.001
30-44	84.1 (19.2)		75.6 (34.8)		66.8 (24.0)		70.9 (19.8)	
45-64	76.1 (23.0)		71.2 (37.8)		64.0 (24.3)		68.8 (20.2)	
Economic class								
A1	97.2 (4.41)	<0.001	83.3 (27.9)	0.003	71.8 (22.3)	0.104	72.3 (18.1)	<0.001
A2	88.3 (14.3)		85.5 (28.0)		70.9 (19.2)		77.4 (16.0)	
B1	82.2 (20.4)		79.9 (32.6)		68.8 (21.8)		76.1 (16.2)	
B2	84.2 (18.7)		75.0 (35.9)		69.7 (24.0)		71.3 (18.6)	
C	79.1 (22.8)		69.8 (36.6)		64.4 (24.8)		68.1 (20.9)	
D	83.6 (18.5)		68.2 (25.3)		72.5 (37.8)		66.9 (22.1)	
Years of Study								
Up to 4	67.8 (27.6)	<0.001	61.9 (42.4)	<0.001	58.1 (29.0)	0.016	62.6 (23.7)	<0.001
5 to 8	76.7 (23.7)		66.1 (38.4)		64.8 (25.8)		64.1 (21.6)	
9 to 11	83.9 (19.3)		74.2 (34.7)		68.2 (24.0)		72.2 (18.3)	
12 or more	86.8 (15.9)		82.4 (30.4)		69.8 (20.6)		75.7 (16.7)	
Chronic disease*								
Yes	74.8 (23.3)	<0.001	64.8 (39.6)	<0.001	58.0 (24.1)	<0.001	64.3 (20.4)	<0.001
No	89.7 (13.9)		84.2 (27.5)		76.8 (19.3)		77.7 (16.2)	
Anxiety*								
Yes	71.5 (24.9)	<0.001	54.6 (40.5)	<0.001	53.8 (24.5)	<0.001	58.3 (21.6)	<0.001
No	85.3 (18.1)		79.9 (31.8)		71.1 (22.1)		74.4 (17.4)	
Depression*								
Yes	66.5 (26.4)	<0.001	51.4 (41.2)	<0.001	52.6 (27.0)	<0.001	54.7 (22.5)	<0.001
No	84.9 (18.1)		78.4 (32.8)		69.9 (22.3)		73.7 (17.6)	

	Vitality		Social Functioning		Role Emotional		Mental Health	
	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P
Gender								
Male	72.4 (16.9)	<0.001	83.6 (19.7)	<0.001	75.0 (35.2)	0.003	77.5 (17.0)	<0.001
Female	63.3 (21.5)		75.0 (25.5)		66.8 (39.5)		71.5 (21.0)	
Age (years)								
20-29	69.2 (18.6)	0.088	81 (21.59)	0.125	71.1 (37.3)	0.334	75.5 (17.9)	0.207
30-44	64.8 (19.7)		76.3 (24.2)		72.4 (36.0)		72.0 (19.9)	
45-64	66.6 (21.6)		77.9 (24.8)		67.8 (39.8)		73.9 (20.6)	
Economic class								
A1	73.8 (15.1)	0.007	84.7 (18.5)	0.004	66.6 (50.0)	0.195	81.5 (15.7)	0.006
A2	72.1 (18.4)		82.6 (20.3)		75.3 (33.9)		79.7 (16.0)	
B1	69.5 (17.9)		83.4 (19.1)		75.0 (33.7)		75.9 (16.9)	
B2	67.7 (20.0)		78.0 (24.2)		69.9 (38.7)		74.3 (20.1)	
C	64.1 (21.3)		74.8 (25.5)		66.0 (39.9)		71.3 (20.7)	
D	64.5 (22.9)		80.1 (25.2)		75.8 (36.5)		71.9 (22.2)	
Years of Study								
Up to 4	60.8 (24)	0.035	78.5 (25.6)	0.194	68.1 (39.1)	0.107	70.3 (21.6)	0.031
5 to 8	63.9 (21.6)		74.7 (26.8)		65.8 (39.6)		70.6 (22.9)	
9 to 11	68.0 (20.9)		78.5 (23.8)		68.4 (39.4)		73.6 (19.9)	
12 or more	68.4 (18.1)		80.0 (21.6)		74.3 (35.5)		76.2 (16.8)	
Chronic disease*								
Yes	60.6 (22.6)	<0.001	70.8 (26.9)	<0.001	61.7 (40.6)	<0.001	67.5 (22.5)	<0.001
No	73.0 (15.7)		85.8 (17.6)		78.5 (33.6)		80.1 (14.1)	
Anxiety*								
Yes	50.1 (22.2)	<0.001	58.2 (26.8)	<0.001	42.1 (39.1)	<0.001	55.4 (22.1)	<0.001
No	71.2 (17.3)		83.6 (19.9)		77.3 (34.2)		78.6 (15.9)	
Depression*								
Yes	44.5 (22.6)	<0.001	50.3 (26.1)	<0.001	37.1 (39.5)	<0.001	49.6 (22.7)	<0.001
No	70.3 (17.6)		82.7 (20.2)		75.2 (35.2)		77.6 (16.2)	

*Self-reported

Table 5 - Percentiles of SF-36 scores on the four domains in the total sample and subsamples grouped by gender and age

	Physical Functioning	Bodily Pain	Role Physical	General Health	Vitality	Social Functioning	Role Emotional	Mental Health
All (n=755)								
5	35	22	0	37	25	25	0	32
10	50	41	0	45	35	50	0	44
25	75	51	50	57	55	62.50	33	64
50	90	72	100	72	70	87.50	100	80
75	95	84	100	87	80	100	100	88
90	100	100	100	95	90	100	100	96
95	100	100	100	97	95	100	100	100
Male (n=290)								
5	50	32	0	37	40	50	0	40
10	65	41	25	47	50	50	0	52
25	83	61	75	62	65	75	66	72
50	95	74	100	77	75	87.50	100	80
75	100	100	100	87	85	100	100	88
90	100	100	100	97	90	100	100	96
95	100	100	100	100	95	100	100	100
Female (n=465)								
5	31.50	22	0	35	20	25	0	32
10	45	31	0	42	30	37.50	0	40
25	70	51	50	57	50	62.50	33	60
50	85	62	100	72	70	75	100	76
75	95	84	100	87	80	100	100	88
90	100	100	100	95	85	100	100	96
95	100	100	100	97	90	100	100	96

	Physical Functioning	Bodily Pain	Role Physical	General Health	Vitality	Social Functioning	Role Emotional	Mental Health
Male Age 20 a 29 (n=102)								
5	66.50	51	25	42.75	45	50	0	40
10	85	54	50	57	50	62.50	0	56
25	90	67	75	67	70	75	66	72
50	100	84	100	81	80	87.50	100	84
75	100	100	100	90	85	100	100	88
90	100	100	100	97	90	100	100	96
95	100	100	100	100	95	100	100	100
Male Age 30 to 44 (n=74)								
5	50	22	0	40.25	38.75	37.50	0	40
10	70	41	10	47	42.50	50	33.33	42
25	80	59	50	62	59	62.50	66	64
50	95	72	100	77	70	87.50	100	76
75	100	84	100	92	80	100	100	84
90	100	100	100	100	87.50	100	100	92
95	100	100	100	100	91.25	100	100	96
Male Age 45 to 64 (n=114)								
5	37.50	31	0	35.25	43.50	50	0	41
10	50	41	12.50	41	50	50	0	52
25	74	51	50	57	62.50	75	66	72
50	87.50	72	100	72	75	87.50	100	84
75	95	84	100	82	85	100	100	92
90	100	100	100	92	95	100	100	96
95	100	100	100	98	100	100	100	100
Female Age 20 to 29 (n=98)								
5	0	65	22	0	37	25	25	0
10	22.50	70	41	22.50	47	35	49	0
25	50	80	51	50	62	50	62.50	33
50	100	95	72	100	77	70	75	83
75	100	100	84	100	87	80	100	100
90	100	100	100	100	95	85	100	100
95	100	100	100	100	97	90	100	100

	Physical Functioning	Bodily Pain	Role Physical	General Health	Vitality	Social Functioning	Role Emotional	Mental Health
Female Age 30 to 44 (n=138)								
5	0	30	22	0	36.90	20	25	0
10	0	54.50	31	0	42	30	37.50	0
25	50	75	51	50	54	50	50	33
50	100	90	62	100	68.50	65	75	100
75	100	95	84	100	82	80	100	100
90	100	100	100	100	92	85	100	100
95	100	100	100	100	97	90	100	100
Female Age 45 to 64 (n= 229)								
5	0	25	22	0	28.15	15	25	0
10	0	35	22	0	42	30	37.50	0
25	25	60	41	25	53	50	62.50	33
50	100	80	62	100	72	70	75	100
75	100	90	84	100	87	80	100	100
90	100	100	100	100	95	85	100	100
95	100	100	100	100	97	90	100	100

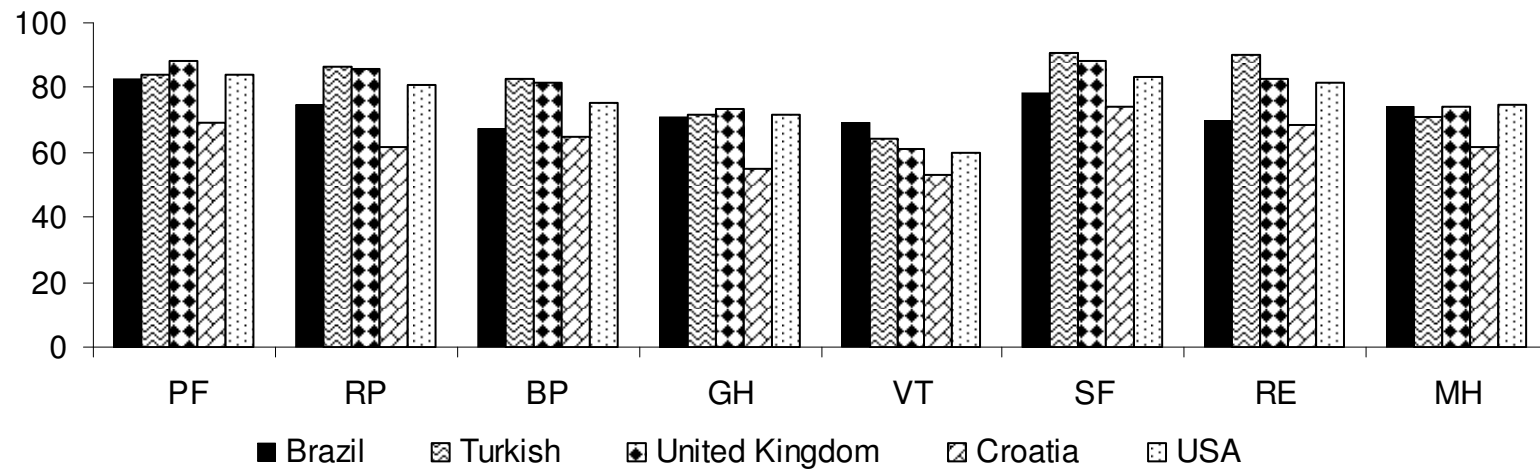


Figure 1 - Means scores of SF-36 domains in the general population of Brazil and other countries

ARTIGO 5

**QUALITY OF LIFE IN BRAZIL: NORMATIVE VALUES
FOR THE WHOQOL-BREF IN A SOUTHERN
GENERAL POPULATION SAMPLE**

QUALITY OF LIFE IN BRAZIL: NORMATIVE VALUES FOR THE WHOQOL-BREF IN A SOUTHERN GENERAL POPULATION SAMPLE

Luciane N. Cruz ¹, Carisi A. Polanczyk^{1,2}, Suzi A. Camey ³,
Juliana F. Hoffmann¹, Marcelo P. Fleck⁴.

Artigo a ser submetido para o periódico Quality of Life Research

Graduate Studies Program in Epidemiology ¹, Statistics Department³, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; Cardiology Division of Hospital de Clinicas de Porto Alegre⁴, Brazil; Graduate Studies Program in Psychiatry², Federal University of Rio Grande do Sul, Porto Alegre, Brazil

Corresponding author:

Luciane Nascimento Cruz

Rua Ramiro Barcelos, 2350/second floor

Porto Alegre – RS – Zip Code 90035-003

Phone: 55 51 3356-7880

Fax: 55 51 3342-7959

E-mail: Incruz@terra.com.br

Word count : 3.197

ABSTRACT

Objective: Normative data for WHOQOL-bref are scarce in the literature and unavailable in Latin American countries. The main objective of this study was to provide normative scores of WHOQOL-bref in a general population sample in Brazil and to describe differences in mean scores according to some sociodemographic characteristics.

Methods: WHOQOL-bref was applied to a randomly selected sample of the general population of Porto Alegre. Participants were literate people aged 20 to 64 years. The questionnaires were self-administered in the presence of an interviewer in the respondent's home.

Results: The response rate was 68% and the final sample contained 751 respondents (38% male, 62% female). Decreased quality of life was observed in females, in the 30-44 year age group, in subjects from a lower economic class, less educated and reporting a chronic medical condition. The mean scores of the WHOQOL-bref and percentiles of scores are reported as normative data for the general population.

Conclusion: Our results can be useful to researchers using the WHOQOL-bref to compare their results with normative data from a randomly selected sample of general population. Additionally, the ability of WHOQOL-bref to discriminate different population subgroups makes it an important tool to identify vulnerable groups in epidemiological surveys.

Keywords: Quality of Life, WHOQOL-bref, Normative Data, Brazil.

RESUMO

Objetivo: Dados normativos para o WHOQOL-bref são escassos na literatura e indisponíveis em países da América Latina. O principal objetivo deste estudo foi o de fornecer escores normativos do WHOQOL-bref em uma amostra da população geral no Brasil e investigar diferenças nos escores em relação a características sociodemográficas.

Métodos: O WHOQOL-bref foi aplicado a uma amostra aleatória da população geral de Porto Alegre. Os participantes eram pessoas alfabetizadas com idade entre 20 a 64 anos. Os questionários foram auto-administrados, na presença de um entrevistador no domicílio dos participantes.

Resultados: A taxa de resposta foi de 68% e a amostra final constituiu-se de 751 respondentes (38% homens, 62% do sexo feminino). Diminuição da qualidade de vida foi observada em mulheres, em indivíduos na faixa etária de 30 a 44 anos, de uma classe econômica mais baixa, com menor nível de escolaridade e relatando uma condição médica crônica. As médias e percentis dos escores do WHOQOL-bref são apresentados em tabelas e gráficos, representando os dados normativos para a população geral.

Conclusão: Os resultados deste estudo podem ser utilizados por pesquisadores que aplicam o WHOQOL-bref em grupos específicos da população para comparações dos escores com os valores de referência populacionais. Além disso, a capacidade demonstrada pelo WHOQOL-bref de discriminar diferentes subgrupos da população o torna uma importante ferramenta para identificar grupos vulneráveis em estudos epidemiológicos.

INTRODUCTION

The quality of life (QOL) measurement has been increasingly used in recent decades as an outcome measure in clinical trials, in effectiveness studies for different interventions, in the health technology assessment, and in epidemiological surveys to assess the subjective health and well-being of the population. The inclusion of this measure in health care is an important supplement to traditional morbidity and mortality indicators, because it takes into account the perception of patients regarding their treatments (1).

In this context, the Quality of Life Group of the World Health Organization (WHOQOL group) has developed an instrument to assess QOL in a cross-cultural perspective for international use (2). The measure has been developed considering the concept of QOL as defined by the World Health Organization: QOL is "the individual's perception of their position in life, in the context of culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns". This definition reflects the view that QOL refers to a subjective assessment embedded in a cultural, social and environmental context (3). An instrument with 100 questions was initially developed, the WHOQOL-100 (4), involving the participation of several countries, representing different cultures, having a Brazilian version being developed (5). However, the need for shorter instruments that require less time for completion, but with satisfactory psychometric characteristics, led the WHOQOL-group to develop an abridged version of WHOQOL-100, the WHOQOL-bref (3).

The Brazilian Portuguese version of the WHOQOL-bref showed satisfactory characteristics of internal consistency, discriminated validity, criterion validity, concurrent validity, and test-retest reliability (6). It was proved to be an useful alternative for situations where the long version is difficult to apply such as epidemiological studies and/or studies using multiple assessment tools (2).

The increasing use of the WHOQOL-bref as an instrument for QOL assessment demands the need for normative data to be used as reference values for

comparisons between groups. Tables of normative scores, taken from surveys of randomly selected subjects from the general population, provide a useful guide to interpret results in the lack of a “gold standard” in quality of life measures (7).

Normative data for WHOQOL-bref are still scarce in the international literature (8;9) and not available in Brazil. In order to fill this gap, the objectives of this study were a) to provide WHOQOL-bref scores in a sample of the general population of a region in Brazil, and b) to describe differences in mean scores according to socio-demographic characteristics of individuals.

METHODOLOGY

Sampling

Target population was consisted of the general population of Porto Alegre, a city with 1,436,123 inhabitants located at South of Brazil. It is the capital of one of the more developed states of the country, with 97% of the population living in urban areas, per capita GDP of approximately US\$ 13,000.00 and a literacy rate of 96.7% (10).

The sample was planned aiming 800 individuals. A two-stage cluster random selection design was used. In a first stage of sampling, a random sample of 108 census sectors of the city was obtained, divided by the Brazilian Institute of Geography and Statistics (IBGE). To calculate the number of households to be visited, the average number of adults per household was considered, and the population's proportion in each one of the strata the study aimed to reach, that is, men and women in the age ranges of 20-29, 30-44, 45-64 years. In each sector, 7 households to be visited were systematically selected and all residents were invited to participate in the study if they met the following inclusion criteria: age ranging from 20 to 64 years; be literate; not having any physical or mental limitation that could prevent the reading and understanding of the instruments. If the residents were not found in the first visit, another two visits in different days and times, including non-business hours, were carried out. A cover letter containing the team identification and purposes of the study, time to be spent on the interview and phone numbers was provided to residents present in the first visit, or deposited in the mailbox of the selected households.

The interviews were conducted in the homes of participants. The interviewer explained how to complete the questionnaires and these were filled by the respondent, except for socio-demographic questionnaire that was completed by the interviewer. The registration of start time and end of filling out each questionnaire as well as checking the material answered by the participant was performed by the team at the end of each interview.

Instruments

The WHOQOL-bref is a generic quality of life instrument composed of 26 questions, two of them measuring overall and general health (not included in the scoring) and the other representing each of the 24 facets that compose the original instrument, the WHOQOL-100. The 24 questions are divided into 4 areas: physical, psychological, social relationships and environment (Appendix 1). Each item scores from 1 to 5 on a Likert scale. The scores are then transformed into a linear scale between 0 and 100, with 0 being the least favorable quality of life and 100 being the most favorable one.

A standardized questionnaire was used to obtain socio-economic and demographic data and contained the following variables: gender, age, race, marital status, practice of any religion, employment status and economic class, number of medical consultations and admissions in the last year, smoking and alcohol use. The presence of chronic diseases was assessed by a list of diseases with dichotomous response (yes/no): hypertension, diabetes, ischemic cardiopathy (infarct/angina), heart failure, arthrosis/arthritis, cerebrovascular accident, chronic bronchitis/emphysema, asthma, kidney disease, cancer, HIV/AIDS, back pain, depression and anxiety, and one open question codified as "others".

The economic class was assessed by an index called Brazil Criterion (Critério Brasil) (11) which divides the population into classes according to their purchasing power and schooling of the family head. The classification and its equivalence concerning approximate mean family income in American dollars would be: Class A1: Average family income of US\$ 3,800; Class A2: US\$ 2,300; Class B1: US\$ 1,400; Class B2: US\$ 800; Class C: US\$ 460; Class D: US\$212, and Class E: US\$ 103.

Statistical analysis

Continuous data are expressed as means \pm standard deviation and categorical in percentage. Comparisons of QOL mean scores among groups according to socio-demographic characteristics were performed by ANOVA, Brown-Fosythe or t-test. ANOVA was used when there was homogeneity of variances and Brown-Fosythe when there was not, both of them for comparisons among two or more groups. Levene's test was used to perform tests of homogeneity of variances.

For all tests, a significance level < 0.05 was established. Data were analyzed using SPSS for Windows, version 13.0 (IBM company, Chicago) and Microsoft Office Excel 2003.

RESULTS

Study population

Data collection was performed from July/2007 to July/2009. Of the 1057 households visited, we could conduct interviews in 514 (49%), because for the others it was impossible to contact dwellers after 3 consecutive visits or people refused to welcome the study team. From 1119 eligible individuals, identified and contacted, 758 participated in the project, achieving a response rate of 68%. The number of eligible individuals includes all the households in which the research team was able to contact, even households where people refused to participate, because we could collect information regarding the number of people aged from 20 to 64 living at the addresses contacted.

A total of 6 exclusions were necessary, 2 for error in recording the age and other 5 (0.7%) for having more than 20% of missing data in the WHOQOL-bref questions, as recommended by the WHOQOL group (12). Of the 751 subjects with data available for analysis, 703 responded to the 26 items of the instrument, leading to a completeness rate of 93%. The items that most frequently had missing values were item 5 ("How much do you enjoy life?") in 9 cases (1.2%), and item 21 ("How satisfied are you with your gender life?") in 11 cases (1.5%). The mean of time spent to complete the WHOQOL-bref was 8 minutes.

Table 1 shows the distribution of the study population compared to the general population. In the sample, there was a slightly lower percentage of men aged 30-44 years, as well as a greater proportion of women aged 46 to 64 years. The subrepresentativity of D economy class (7% compared to 28% in the general population) and the absence of participants in the E class was an expected finding, since one of the exclusion criteria of the study was illiteracy, which is generally more prevalent in these population strata. In addition, 8 (6.7%) census sectors had to be excluded from the sample because they are places with high crime rates, which could compromise the safety of team members. Such exclusions, because they involve neighborhoods possibly inhabited by people of lower purchasing power, may also justify the low level of D class and the lack of E class.

The socio-demographic characteristics of the sample are presented in Table 2. There was a higher percentage of female individuals, married, white, practicing a religion. We also observed a higher number of participants over 12 years of study, in C Class and with a formal employment. The mean years of schooling of the sample was 11 years.

Regarding self-reported illnesses, 368 (49%) reported having some physical and/or mental disease. 103 (14%) of the interviewed said they had depression, and 156 (21%) anxiety. Among the chronic physical conditions in the questionnaire, the most frequently reported were: hypertension (13%), arthritis (8%), asthma (8%), and diabetes (4%).

Quality of Life

The means in each domain for the 751 respondents are described in Table 3. The highest mean was found in the social domain (76.2) and the lowest in the physical domain (58.9).

There was a difference in the mean scores of quality of life in different strata of the sample in relation to age and gender, as can be seen in Table 3. Women had worse scores in all domains, but in social domain the difference was not statistically significant. The age group of 30 to 44 years had lower means compared to those of younger and older in all domains, but the difference was statistically significant only in the psychological and social areas.

The mean scores also varied according to economic class and educational level. Higher values were recorded for the group belonging to the A1 class, with a gradual reduction to the D class in all domains. The QOL scores also decreased in direct proportion to education, except in the social domain where the group with 5 to 8 years of study was the only one to have lower mean score in relation to groups with up to 4 years, 9 to 11 years, and over 12 years of study.

The fact of being identified as having a chronic illness, mental or physical, seems to lead to less satisfaction of individuals in relation to all aspects of quality of life measured by the WHOQOL-bref. Differences in mean scores between individuals who reported no illness and those who did was statistically significant ($p < 0.001$) in almost all domains, with the lower values for the second group.

Table 4 shows the scores in each domain of the WHOQOL-bref in percentiles 5, 10, 25, 50, 75, 90, 95 for the total sample and for each group according to gender and age. Figure 1 shows curves for the percentiles 25, 50 and 75 of scores for men and women by age, in each domain of WHOQOL-bref. The description in percentiles is to make the use of scores reported more practical for future comparisons. For example: it might more clear the information that the QOL score of a particular patient or group of patients is the 50th percentile of the general population scores than the information that the QOL is "better" "worse" or "equal".

DISCUSSION

This study provides scores of quality of life measured by the WHOQOL-bref based on a sample of the general population of a southern Brazilian city. These data can be used as reference for comparison of groups of individuals in different clinical situations. In the absence of a "gold standard" comparative in QOL, literature recommends that normative values obtained in the general population serve as parameter (7;13).

The WHOQOL-bref proved to be an instrument easy to apply with high acceptability, since 93% of participants answered all 26 questions.

The description of quality of life through a multidimensional instrument, which also focuses on the lives of individuals unrelated to health, seems to reflect the socio-economic and cultural context where people are located. Comparing the results of

this study to those found in studies of normatization of the WHOQOL-bref in developed countries such as Netherlands, for example, it was observed that in Physical and Environment domains our means were lower than the means seen in the general Dutch population (14). These domains measure satisfaction with health, dependence on medication, safety, financial resources, access to information, recreation and transportation. On the other hand, in the domains assessing aspects of social life, personal relationships and social support, the scores found in the study sample was higher in relation to the Dutch population scores. The comparison of scores among economic classes indicates that individuals of lower classes are less satisfied than those of higher class in all domains measured by the WHOQOL. The relationship among objective economic measures such as per capita GDP and subjective well-being is controversial. Kenny (15) aggregated results of different studies that examined this relationship and concluded that there is weak evidence of a positive association between economic power and satisfaction in developing countries. The literature suggests that social questions of poverty might have a much greater impact on subjective well-being than income per se, for example, the fact that they feel marginalized for not owning property acquired by their peers in social groups. The criterion used in the present study for division into economic classes allows for the possession of consumer goods and not the individual's income, so the purchasing power of the lower strata is less. Lower power consumption implies in our society less access to leisure, information, education, health and safety, which accounts for the lower satisfaction with physical, psychological, social and environmental aspects. This information has been corroborated by recent studies which have shown that subjective measures of well-being contain genuine information about the objective quality of life of a population. Oswald et al. (16) compared the data from a large study that measured the objective quality of life through economic, social and environmental indicators with data from the U.S. Behavioral Risk Factor Surveillance System, CDC, which assessed life satisfaction in a sample of 1.3 million of Americans, through a subjective question: "In general, how satisfied are you with your life"? The authors found a statistically significant correlation with a coefficient of Pearson correlation of 0.6, considered high for the behavioral sciences.

The observed difference in QOL scores according to demographic variables emphasizes the need for using the standards for each subgroup for comparison purposes. Women had worse scores in nearly all domains. A similar result was observed in the population of Kuwaiti (9) and other studies that applied the WHOQOL-bref in different populations (12). Concerning age, the age group of 30 to 44 years had lower score means compared to those of younger and older, with a statistically significant difference in the psychological and social areas. This finding is controversial through studies, since some authors found a QOL decrease proportional to the increase of age (9), while others have also noticed a better quality of life for people above 45 years in the psychological, social and environmental areas (12). Skevington states that younger people consider themselves more independent than older people, so that age might be related to how people evaluate their own quality of life (12). Another possibility is that older individuals are better adapted to the limitations imposed by aging, with greater maturity to tolerate frustrations, showing to be satisfied with their life and the goals achieved than the younger people, less resigned and still being more idealistic.

The WHOQOL-bref was able to differentiate people with physical and mental illnesses from those who considered themselves healthy. The means scores of those reported with some disease was significantly lower than those of non-patients in all areas. The differences were most evident between the groups with and without self-reported depression and anxiety. This ability of WHOQOL-bref to distinguish patients with emotional distress from those healthy had already been seen in Brazilian populations of patients with chronic physical diseases (17;18) and primary care patients (19;20), and the present study emphasizes these findings adding information related to this property of the instrument also in the general population.

One aspect that should be highlighted is the careful interpretation of the scores for different subgroups. These are cross-sectional data that describe the quality of life of individuals in a single point in time. It does not mean that changes in people's lives that cause them to be part of another group, for example, get older, with higher socio-economic level and more years of study, will result in an improved quality of life. For these changes to be detected, the assessments should be made in follow-up studies of the population.

One of the caveats of this study is the fact that the sample had a subrepresentativity of the general population in relation to the lower socioeconomic classes. We recognize that some criteria followed by the protocol such as to exclude illiterate and preserve the team to work on areas of greatest risk for urban violence may have led to a smaller representation of D and E classes. Since the quality of life scores decreased progressively in lower socioeconomic classes, one can infer that the classes D and E would score even lower. Then, using the results presented in this paper to compare quality of life between groups, researchers should be aware that for individuals pertaining to classes D and E, values are probably overestimated.

The data here presented should be used with caution regarding the extrapolation of results for the country as a whole. Brazil has 5 regions in very different socio-economic, geographical and cultural terms. Thus, we believe that normative scores of quality of life should be described by each region so that local characteristics are considered. Replication of this study in other regions could contribute to the achievement of national values. In Brazilian populations with a profile similar to that found in the population of the city of Porto Alegre, the reported data can be used for comparisons.

While we recognize some caveats of the study, it is important to emphasize the difficulty of conducting a population survey in our country. The high rates of violence in urban centers leads to a fear on the part of society to welcome strangers into your home, even after being identified as members of a research team. In many census sectors, people use to live in buildings with security systems that greatly hinder access to residents. For these reasons, it was necessary to adopt the strategy of replacement of losses and refusals, visiting a number of households larger than planned in order to obtain the required number of interviews.

One of the major strengths of this study for future research is to provide tables of the WHOQOL-bref scores to be used as a comparative standard for quality of life assessments in groups of patients with different diseases. This research was undertaken according to international methodological recommendations for normatization of QOL instruments, using a random sample of the general population, describing QOL scores for different age and gender strata, and response rate over two thirds (13).

The WHOQOL-bref proved to be a sensitive measure for socio-demographic

variables, being a useful tool for identifying vulnerable groups and describing the profile of quality of life of the population in its different strata. Since it is easy to apply, this measure could be included in studies where it is not possible to add more instruments and it can be used as an indicator of the presence of physical and psychiatric conditions such as depression and anxiety.

This instrument can be a useful tool for public health as an outcome measure to assess the satisfaction level of users of health services, establish which areas of patients' lives are being affected by the interventions and evaluate changes in quality of life after different treatments.

Acknowledgements: We thank the Research Incentive and Event Fund of Hospital de Clínicas de Porto Alegre for the financial aid in translating this article. Dr. Luciane Cruz received graduate research scholarship from CAPES, Brazil. Prof. Polanczyk and Prof. Fleck received a research scholarship from CNPq/Brazil

This study was funded by CNPQ/Brazil (Edital MCT-CNPq / MS-SCTIE-DECIT – Nº 36/2005).

References

- (1) Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995 Jan 4;273(1):59-65.
- (2) Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, et al. [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-brief]. *Rev Saude Publica* 2000 Apr;34(2):178-83.
- (3) Development of the World Health Organization WHOQOL-BRIEF quality of life assessment. The WHOQOL Group. *Psychol Med* 1998 May;28(3):551-8.
- (4) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 1998 Jun;46(12):1569-85.
- (5) Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, et al. [Application of the Portuguese version of the instrument for the assessment of quality of life of the World Health Organization (WHOQOL-100)]. *Rev Saude Publica* 1999 Apr;33(2):198-205.
- (6) Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, et al. [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref]. *Rev Saude Publica* 2000 Apr;34(2):178-83.
- (7) Fayers PM, Machin D. *Quality of Life. The Assessment, analysis and interpretation of patient-reported outcomes.* Second ed. West Sussex, England: John Wiley & Sons; 2007.
- (8) Noerholm V, Groenvold M, Watt T, Bjorner JB, Rasmussen NA, Bech P. Quality of life in the Danish general population--normative data and validity of WHOQOL-BREF using Rasch and item response theory models. *Qual Life Res* 2004 Mar;13(2):531-40.
- (9) Ohaeri JU, Awadalla AW, Gado OM. Subjective quality of life in a nationwide sample of Kuwaiti subjects using the short version of the WHO quality of life instrument. *Soc Psychiatry Psychiatr Epidemiol* 2009 Aug;44(8):693-701.
- (10) IBGE Cidades. Instituto Brasileiro de Geografia e Estatística (IBGE) Available from: URL: www.ibge.gov.br. Last update 12.02.10
- (11) Critério Econômico Brasil 2003. Associação Brasileira de Empresas de Pesquisa. Available from: URL: www.abep.org/novo/default.aspx. Last update 12.02.10
- (12) Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BRIEF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res* 2004 Mar;13(2):299-310.

- (13) Gandek B, Ware JE, Jr. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998 Nov;51(11):953-9.
- (14) Noerholm V, Groenvold M, Watt T, Bjorner JB, Rasmussen NA, Bech P. Quality of life in the Danish general population--normative data and validity of WHOQOL-BRIEF using Rasch and item response theory models. *Qual Life Res* 2004 Mar;13(2):531-40.
- (15) Kenny C. Does development make you happy? Subjective wellbeing and economic growth in developing countries. *Social Indicators Research* 2005;73:199-219.
- (16) Oswald AJ, Wu S. Objective confirmation of subjective measures of human well-being: evidence from the U.S.A. *Science* 2010 Jan 29;327(5965):576-9.
- (17) Cruz LN, de Almeida Fleck MP, Polanczyk CA. Depression as a determinant of quality of life in patients with chronic disease: data from Brazil. *Soc Psychiatry Psychiatr Epidemiol* 2009 Sep 22.[Epub ahead of print]
- (18) Rabin EG, Heldt E, Hirakata VN, Fleck MP. Quality of life predictors in breast cancer women. *Eur J Oncol Nurs* 2008 Feb;12(1):53-7.
- (19) Fleck MP, Lima AF, Louzada S, Schestasky G, Henriques A, Borges VR, et al. [Association of depressive symptoms and social functioning in primary care service, Brazil]. *Rev Saude Publica* 2002 Aug;36(4):431-8.
- (20) da Silva Lima AF, de Almeida Fleck MP. Subsyndromal depression: an impact on quality of life? *J Affect Disord* 2007 Jun;100(1-3):163-9.

Table 1 - Socio-demographic comparison between total sample and the general population of Porto Alegre

	Total sample (n=751) n (%)	General Population * (%)
Men		
20-29 years	102 (14)	15
30-44 years	74 (10)	16
45-64 years	112 (15)	15
Women		
20-29 years	97 (13)	15
30-44 years	138 (18)	19
45-64 years	228 (30)	20
Economic class **		
A1	9 (1)	1
A2	77 (10)	5
B1	116 (15)	7
B2	189 (25)	17
C	307 (41)	38
D	52 (7)	28
E		5

*Source: IBGE: WWW.ibge.org.br ** Source: Brazil Criterion - www.abep.org

Table 2 - Socio-demographic characteristics of the study population

Variables	N (Total=751)	%
Age (mean - SD)	41 (13)	
Gender		
Male	288	38
Female	463	62
Race		
White	606	81
Black	73	10
Brown	53	7
Others	19	2
Religion *		
Yes	402	54
No	348	46
Marital status		
Married	445	59
Not married	306	41
Years of study*		
Up to 4 years	46	6
5 to 8 years	167	22
9 to 11 years	257	34
12 years or more	277	37
Job status (%)		
Employed	378	50
Informal job	62	8
Unemployed	50	7
Housewife	115	15
Student	42	6
Retired	69	9
Other	35	5

*Variables with missing data

Table 3 - Mean scores (SD) of WHOQOL-bref in the total population sample and subsamples stratified by gender, age, economic class, years of study and presence of a chronic disease

	Physical		Psychological		Social		Environment	
	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P	Mean (SD)	P
All subjects	58.9 (10.5)		65.9 (10.8)		76.2 (18.8)		59.9(14.9)	
Gender								
Men	60.3 (9.9)	0.001	65.9 (10.8)	<0.001	73.5 (18.4)	0.262	62.9 (13.8)	0.003
Women	57.8 (10.5)		62.2 (12.5)		71.9 (19.7)		59.7 (16.2)	
Age (years)								
20-29	58.9 (10.5)	0.360	65.9 (10.7)	0.006	76.2 (18.8)	0.001	59.9 (14.9)	0.109
30-44	57.9 (10.5)		62.5 (12.4)		68.9 (21.7)		59.7 (15.1)	
45-64	59.2 (10.2)		63.0 (12.3)		72.7 (17.3)		62.2 (15.9)	
Economic class								
A1	64.7 (7.46)	<0.001	71.3 (7.6)	0.001	77.8 (9.32)	0.045	76.4 (7.8)	<0.001
A2	62.3 (9.4)		66.6 (11.5)		76.6 (17.5)		72.9 (13.5)	
B1	60.0 (9.9)		64.7 (11.4)		74.8 (17.5)		67.9 (12.5)	
B2	60.1 (9.5)		64.7 (10.5)		73.1 (17.5)		60.9 (14.2)	
C	56.9 (10.8)		62.1 (13.0)		70.5 (20.8)		55.9 (14.6)	
D	55.8 (10.8)		61.4 (11.6)		70.9 (21.5)		54.0 (16.5)	
Years of Study								
Up to 4 years	55.1 (10.26)	0.014	59.5 (13.0)	0.002	73.2 (17.5)	0.048	53.7 (16.0)	0.000
5 to 8 years	57.8 (11.6)		61.3 (12.9)		68.9 (22.7)		56.7 (15.6)	
9 to 11 years	58.8 (9.9)		64.4(12.4)		73.9 (17.9)		57.8 (14.3)	
12 years or more	59.9 (9.9)		65.0 (10.6)		73.4 (18.2)		67.4 (13.9)	
Self-Reported Anxiety								
Yes	54.5 (10.6)	<0.001	55.9 (14.0)	<0.001	63.0 (21.3)	<0.001	55.2 (16.2)	<0.001
No	59.9 (10.03)		65.7 (10.5)		75.0 (17.8)		62.4 (14.9)	
Self-Reported Depression								
Yes	51.9 (11.2)	<0.001	52.5 (13.7)	<0.001	58.01 (20.5)	<0.001	52.8 (17.0)	<0.001
No	59.9 (9.8)		65.4 (10.7)		74.8 (17.9)		62.2 (14.8)	
Self-Reported Chronic Disease								
Yes	57.5 (11.1)	<0.001	61.4 (13.0)	<0.001	69.3 (20.5)	<0.001	59.3 (16.4)	0.004
No	60.2 (9.4)		65.9 (10.4)		75.8 (17.4)		62.5 (14.2)	

Table 4 - Percentiles of mean scores of WHOQOL-bref in the total sample and subgroups divided by sex and age

	Physical	Psychological	Social	Environment
All (n=751)				
5	39	42	42	34
10	46	46	50	41
25	54	58	67	50
50	61	67	75	63
75	64	71	83	72
90	71	79	92	81
95	75	79	100	84
Males (n=288)				
5	43	42	42	38
10	46	50	50	44
25	54	63	67	53
50	61	67	75	63
75	68	75	83	72
90	71	79	92	81
95	75	79	100	84
Females (n=463)				
5	39	38	33	31
10	43	46	50	38
25	54	54	58	50
50	61	63	75	59
75	64	71	83	72
90	71	75	97	81
95	75	80	100	84
Males Age 20 to 29 (n=102)				
5	40	46	35	38
10	46	51	50	44
25	54	63	67	50
50	61	67	75	63
75	68	75	92	72
90	71	79	100	78
95	71	83	100	84
Males Age 30 to 44 (n=74)				
5	43	41	25	38
10	46	42	38	45
25	54	58	58	53
50	61	67	71	63
75	65	71	83	69
90	71	75	92	80
95	76	79	94	84
Males Age 45 to 64 (n=112)				
5	45	48	42	38
10	54	54	58	47
25	57	63	67	56
50	61	67	75	66
75	68	75	83	75
90	71	79	92	81
95	79	79	100	88

	Physical	Psychological	Social	Environment
Females Age 20 to 29 (n=97)				
5	39	46	42	31
10	46	53	48	38
25	50	58	67	47
50	61	67	75	59
75	66	71	92	69
90	71	79	100	79
95	75	83	100	83
Females Age 30 to 44 (n=138)				
5	37	38	25	31
10	41	42	42	38
25	50	54	58	47
50	59	63	75	59
75	64	71	83	72
90	71	75	100	81
95	75	83	100	84
Females Age 45 to 64 (n= 228)				
5	39	38	42	33
10	43	44	50	41
25	54	50	58	50
50	61	63	75	63
75	64	71	83	72
90	71	75	92	82
95	75	79	100	88

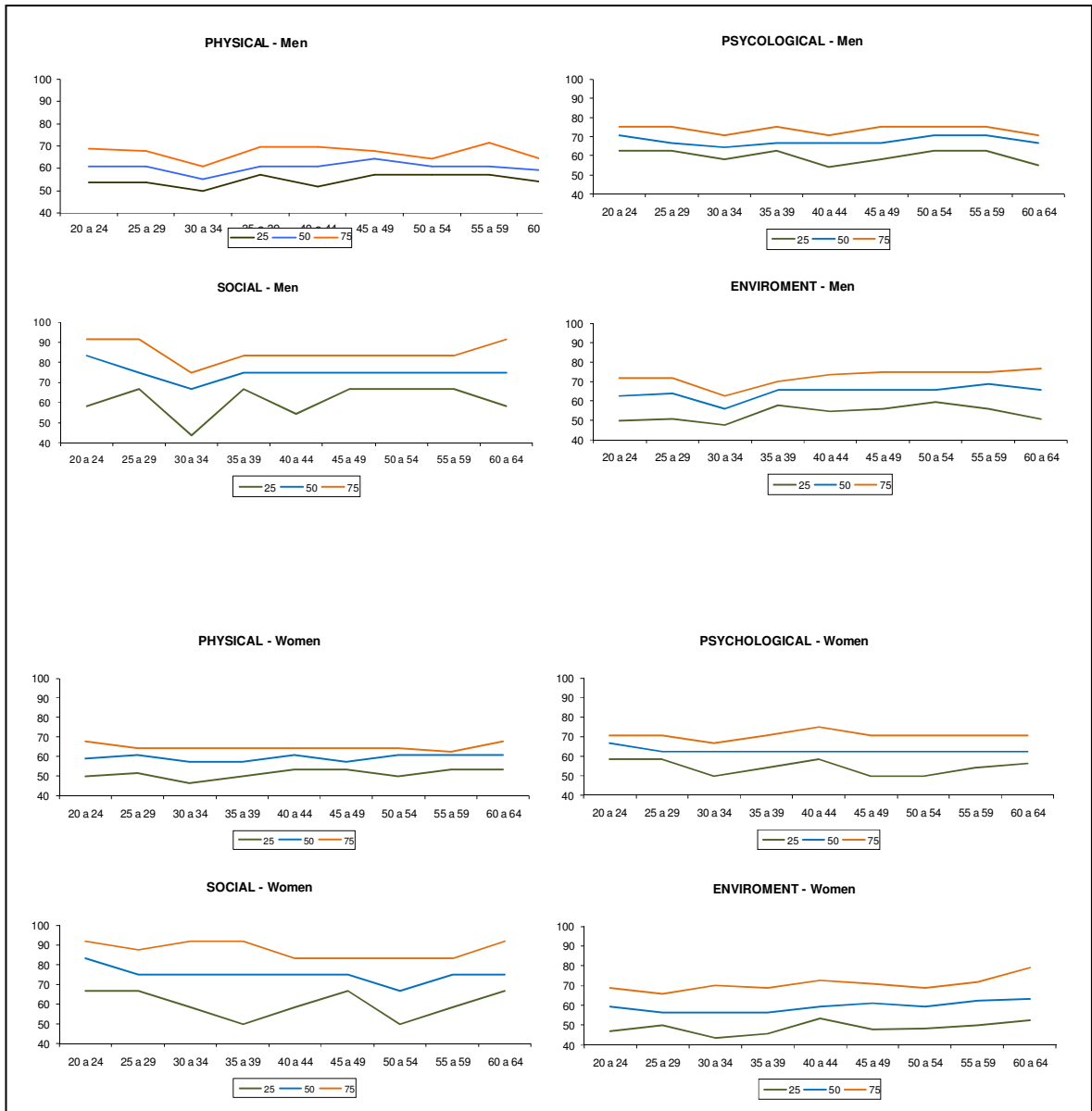


Figure 1 - Percentiles charts of the scores of each domain of WHOQOL-bref for men and women

Appendix 1

Domains and facets within domains of WHOQOL-bref

Domain 1 – Physical health	
9.	Pain and discomfort
10.	Energy and fatigue
11.	Sleep and rest
12.	Mobility
13.	Activities of daily living
14.	Dependence on medical substances and medical aids
12.	Work capacity
Domain 2 – Psychological	
12.	Positive feelings
13.	Thinking, learning, memory and concentration
14.	Self-esteem
15.	Bodily image and appearance
16.	Negative feelings
24.	Spirituality/religion/personal beliefs
Domain 3 – Social relationships	
15.	Personal relationships
16.	Social support
15.	Sexual activity
Domain 4 – Environment	
23.	Freedom, physical safety and security
24.	Home environment
25.	Financial resources
26.	Health and social care: accessibility and quality
27.	Opportunities for acquiring new information and skills
18.	Participation in and opportunities for recreation/leisure activity
22.	Physical environment (pollution/noise/traffic/climate)
23.	Transport

WHOQOL-group (3)

6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

Os resultados da presente tese disponibilizam medidas de qualidade de vida que poderão ser utilizadas como desfecho em estudos de custo-efetividade, estudos clínicos e epidemiológicos.

Os instrumentos de avaliação de qualidade de vida adequados para uso em análises econômicas, as medidas baseadas em preferências ou medidas de utilidade, não haviam sido adaptadas para uso no Brasil até o momento. Técnicas para obtenção de índices de utilidade como o *standard gamble*, são complexas e intuitivamente poderiam ser consideradas inadequadas para aplicação na população geral em nosso meio. Entretanto, os resultados do estudo de medida de preferências na população geral de Porto Alegre demonstraram que a porcentagem de respostas inconsistentes foi semelhante à encontrada em países desenvolvidos. Além disso, uma maior dificuldade em responder a este tipo de questionário foi encontrada por um subgrupo específico de pessoas com menor nível educacional e de maior idade, ressaltando a necessidade do uso de outro tipo de medida de utilidade neste estrato populacional. Alternativas a esta limitação do instrumento para este subgrupo pode ser encontrada nos resultados do estudo que estimou o modelo para obtenção dos índices de utilidade através do instrumento SF-6D. Utilizando esta estratégia, substitui-se a aplicação de uma técnica complexa e de alto custo para aplicação como o *standard gamble* por outro método mais simples e rápido, a aplicação do SF-36. As respostas ao SF-36 poderão ser transformadas em valores de utilidade através do algoritmo SF-6D. O modelo recomendado desempenhou-se tão bem quanto o modelo do estudo original britânico do SF-6D em termos de habilidade preditiva, mas difere no tipo de modelagem e no valor dos coeficientes, justificando sua adaptação para contexto nacional.

Os resultados dos estudos de normatização dos instrumentos psicométricos SF-36 e WHOQOL-breve apresentam uma alternativa ao fato de não haver “padrão-ouro” para medidas de qualidade de vida. Os valores normativos dos escores descritos para cada estrato de acordo com sexo e faixa etária poderão ser utilizados como referência pelo fato de terem sido coletados em amostra aleatória da população geral. Os dados obtidos da aplicação destes instrumentos em grupos

específicos de indivíduos podem ser mais bem compreendidos quando comparados aos dados “normais”.

Tanto o SF-36 como o WHOQOL-breve demonstraram ser boas medidas para avaliação de estado de saúde e de qualidade de vida da população geral, discriminando grupos de acordo com características sociodemográficas.

Em suma, os dados aqui apresentados contribuem para a pesquisa em saúde no contexto nacional através de: a) Descrição de uma metodologia ainda pouco conhecida no Brasil. Pesquisadores interessados neste campo de pesquisa poderão replicar os passos aqui delineados para validação de outros instrumentos no contexto brasileiro. Além disso, a melhor compreensão da metodologia destas medidas facilita e melhora a interpretação de seus resultados por parte dos usuários; b) Disponibilização de uma fórmula estatística, o algoritmo SF-6D, para obtenção de índices de utilidade a partir do SF-36. Estes valores poderão ser utilizados para cálculo de QALYs (*Quality Adjusted Life Years*), as medidas de desfecho recomendadas internacionalmente para estudos de custo-efetividade; c) Possibilidade de obtenção de índices de utilidade na população geral de Porto Alegre através da aplicação do algoritmo SF-6D para os dados obtidos no estudo de normatização do SF-36; d) Apresentação de dados de normatização do SF-36 e WHOQOL-breve, descrevendo parâmetros úteis para estudos comparativos.

7 ANEXOS



ANEXO 1

PROJETO DE PESQUISA

MEDIDAS DE QUALIDADE DE VIDA E UTILIDADE EM UMA AMOSTRA DA POPULAÇÃO BRASILEIRA

1 IDENTIFICAÇÃO

1.1 Título

Medidas de qualidade de vida e utilidade em uma amostra da população brasileira

1.2 Autores

Aluno Pós-Graduando:

Luciane Nascimento Cruz

Psiquiatra

Mestre em Epidemiologia pela Universidade Federal do Rio Grande do Sul

Doutoranda do Programa de Pós-Graduação em Medicina – Epidemiologia, Universidade Federal do Rio Grande do Sul

Orientador:

Carisi Anne Polanczyk

Cardiologista

Doutora em Medicina (Cardiologia) pela Universidade Federal do Rio Grande do Sul

Professora dos Programas de Pós-Graduação em Medicina – Epidemiologia e Cardiologia - Universidade Federal do Rio Grande do Sul

Co-orientador:

Marcelo Pio de Almeida Fleck

Psiquiatra

Doutor em Psiquiatria

Professor Coordenador do Programa de Pós-Graduação em Psiquiatria da Universidade Federal do Rio Grande do Sul

Colaboração

Prof. John Edward Brazier

School of Health and Related Research (SChARR)

Universidade de Sheffield

Inglaterra

1.3 Local de Origem

Programa de Pós-Graduação em Medicina – Epidemiologia

Faculdade de Medicina

Universidade Federal do Rio Grande do Sul

1.4 Local de realização

Laboratório de pesquisa cardiovascular (Cardiolab)
Centro de Pesquisa Experimental
Hospital de Clínicas de Porto Alegre

2 INTRODUÇÃO

A avaliação de qualidade de vida relacionada à saúde vem se tornando nas últimas décadas uma importante medida de desfecho utilizada em ensaios clínicos e em estudos observacionais, medindo o impacto de diferentes doenças e seu respectivo tratamento na vida dos indivíduos e também como medida de efetividade em análises econômicas. A qualidade de vida relacionada à saúde é definida como a percepção da pessoa de sua posição na vida no contexto de sua cultura e sistema de valores nos quais ela vive e em relação aos seus objetivos, expectativas, padrões e preocupações (1). É um conceito multidimensional que incorpora o estado funcional do indivíduo e sua percepção em relação a sua própria saúde.

Os instrumentos para avaliar qualidade de vida podem ser divididos em dois grupos: a) medidas descritivas, ou psicométricas, que geram perfis de saúde em múltiplas dimensões como o estado funcional, estado psicológico e relações sociais; b) métodos baseados em preferências, que geram *utilidade*, um único índice que sintetiza e descreve o valor ou preferência de determinado perfil ou estado de saúde para um indivíduo (2). Esses instrumentos são apropriados para uso em estudos de custo-efetividade (custo-utilidade). Entretanto, os métodos usualmente aceitos para a mensuração de utilidade – *standard gamble*, *time trade-off* – são de difícil entendimento e aplicação em estudos clínicos. Atualmente, instrumentos que agregam estes dois conceitos têm sido desenvolvidos e validados, a fim de que as medidas psicométricas, que geram informação descritiva de um estado de saúde, possam ser também utilizadas em análises econômicas (3).

No Brasil não há instrumentos desenvolvidos para a avaliação de qualidade de vida baseados em preferências, que gerem índice de utilidade, já validados para a população brasileira. Considerando as recomendações internacionalmente estabelecidas para a realização de estudos de análise econômica em saúde (4-6), que ressaltam a necessidade de utilização de parâmetros de utilidade para a análise de custo-efetividade, a carência de metodologias de avaliação da qualidade de vida baseados em preferências determina um grande hiato para a pesquisa em saúde no país. Dessa forma, há necessidade de validação destas medidas para a população brasileira. Portanto, a validação de instrumentos que gerem índices de utilidade para a população brasileira torna-se tarefa imprescindível no desenvolvimento e implantação de metodologias de estudos em economia em saúde no Brasil.

3 OBJETIVOS

3.1 Objetivos principais

1. Adaptação cultural, em uma amostra da população brasileira, da medida de qualidade de vida baseada em preferências SF-6D, originalmente desenvolvida no Reino Unido.

2. Aplicação em amostra da população geral dos instrumentos genéricos de qualidade de vida WHOQOL-breve e SF-36 a fim de disponibilizar escores normativos destes instrumentos.

3.2 Objetivos secundários

1. Descrever a metodologia para medidas de preferências (índices de utilidade), uma vez que esta é uma técnica ainda pouco conhecida no cenário nacional.

2. Avaliar o desempenho da aplicação de técnicas complexas como as utilizadas para medir preferências na população de estudo e as associações dos resultados com características sociodemográficas.

3. Estimar preferências (utilidade) para diferentes estados de saúde em amostra da população brasileira através da aplicação de questionário *standard gamble*.

4 MATERIAL E MÉTODOS

Este é um estudo transversal formado por dois componentes principais:

Componente 1: Aplicação de um método para medir qualidade de vida baseado em preferências para realizar a adaptação cultural para o Brasil do instrumento Short-Form 6D, originalmente desenvolvido no Reino Unido.

Componente 2: Aplicação dos instrumentos genéricos de medida de qualidade de vida WHOQOL-breve e SF-36

Componente 1:

Esta etapa do projeto será realizada em parceria com pesquisadores da Universidade de Sheffield, Reino Unido, com intercâmbio de membros da equipe da UFRGS para planejamento e treinamento da aplicação dos instrumentos. O protocolo será o mesmo utilizado no Reino Unido (7), que constitui-se dos seguintes passos:

- a. Derivação de itens do SF-36 para gerar o sistema descritivo do SF-6D
- b. Seleção de estados de saúde gerados pelo sistema descritivo do SF-6D para serem avaliados pelos indivíduos da amostra
- c. Medida das preferências dos participantes do estudo pelos estados de saúde obtidos no passo anterior através da técnica *standard gamble*
- d. Construção de um modelo estatístico utilizando os valores medidos na etapa c. Esta equação deve ser capaz de extrapolar os valores de utilidade para todos os outros estados de saúde gerados pelo SF-6D, sem a necessidade de que todos sejam avaliados.

Derivação do SF-6D a partir do SF-36

O SF-36 é um instrumento que mede estado de saúde em 8 domínios: capacidade funcional, aspectos físicos, dor, estado geral de saúde, vitalidade, aspectos sociais, aspectos emocionais e saúde mental (8). Cada domínio é formado por itens que questionam o indivíduo sobre sua percepção a respeito de sua própria saúde. O domínio capacidade funcional, por exemplo, tem 10 itens para os quais o respondente tem 3 opções de resposta: “Sim, dificulta muito”, “Sim, dificulta um pouco” e “Não. Não dificulta de modo algum”. As respostas são codificadas de 1 a 3, respectivamente, e a soma destes escores são depois transformadas em uma escala de 0 a 100. Porém, os escores entre os domínios não são comparáveis, pois os domínios não têm o mesmo número de níveis de respostas. Também não há uma relação ordinal entre os itens, isto é, ter um pouco de dificuldade para levantar ou carregar mantimentos tem o mesmo valor de ter um pouco de dificuldade para andar vários quarteirões. Essa estrutura torna o SF-36 um instrumento inadequado para gerar índices de utilidade, pois não seria possível combinar os escores de todos os domínios para formar um único índice, como é exigido para as medidas baseadas em preferências. Assim, o SF-36 será reduzido em tamanho e complexidade para que seja possível aplicar métodos para gerar preferências.

A derivação de itens a partir do SF-36 para gerar o SF-6D seguirá rigorosamente a mesma metodologia do protocolo britânico, utilizando a versão em português do SF-36 (9). Os autores britânicos basearam o processo de seleção de itens nas análises fatoriais realizadas pelos autores do SF-36 original (8), estabelecendo a importância de cada item para os escores globais de cada domínio

Estados de saúde

Um estado de saúde gerado por um sistema descritivo de um instrumento é obtido pela seleção de um nível de cada domínio, obedecendo à ordem dos domínios. Estes estados de saúde são denominados na literatura de estados “multiatributos” (com múltiplos domínios), uma vez que descrevem saúde em vários domínios. Depois de gerar o sistema descritivo do SF-6D a partir do SF-36 conforme descrição acima, estados de saúde podem ser obtidos através da seleção de itens do SF-6D. Uma amostra destes estados é selecionada para realizar o procedimento de avaliação através do *standard gamble*. Serão utilizados os mesmos estados de saúde avaliados no Reino Unido para permitir comparações.

Avaliação dos estados de saúde

A avaliação dos estados de saúde iniciará com um ordenamento de 8 estados, conforme as preferências do entrevistado, do melhor para o pior: 5 estados intermediários gerados pelo SF-6D, mais os estados extremos, o melhor estado, o pior estado e morte imediata. Para esta tarefa, utilizaremos material gráfico, para que os entrevistados tenham uma referência para ordenar os estados conforme sua preferência. Será utilizada uma escala chamada “Termômetro” onde 10 representa o estado mais preferido e 0 o menos preferido. Os estados de saúde serão representados em forma de cartões que serão confeccionados todos na cor branca para evitar que as cores possam influenciar as preferências.

O próximo passo será a aplicação do *standard gamble*, instrumento para medida de preferências baseado em teorias econômicas (10) e por isto considerado como o instrumento “padrão” para este propósito. Esta técnica inicia apresentando ao entrevistado uma descrição de um estado de saúde, que aqui será representada pelos estados de saúde gerados pelo SF-6D. Após ler esta descrição, o entrevistado é orientado a imaginar a situação hipotética na qual ele (ela) é confrontado com uma escolha. As opções disponíveis são continuar a viver no estado descrito em um dos estados intermediários, ou optar por um *gamble* (ou “aposta”). Esta “aposta”, representada pela hipótese de realizar um tratamento, tem 2 possíveis desfechos. O melhor desfecho é o tratamento ter sucesso e ocorrer recuperação imediata da saúde perfeita, com probabilidade p de ocorrer. O pior desfecho é falha no tratamento levando ao pior estado de saúde do SF-6D, com probabilidade $1-p$ de ocorrência. Estas probabilidades variam de forma sistemática durante o exercício, até que o entrevistado seja indiferente entre manter-se no mesmo estado intermediário ou aceitar a “aposta”. Neste ponto, o valor da probabilidade do melhor estado é o valor do índice de utilidade do estado intermediário. O material a ser utilizado foi criado por uma equipe de pesquisadores da Universidade de Mac Master, no Canadá (11) e será reproduzido aqui no Brasil com autorização dos autores para uso neste estudo e está representado na Figura 1.

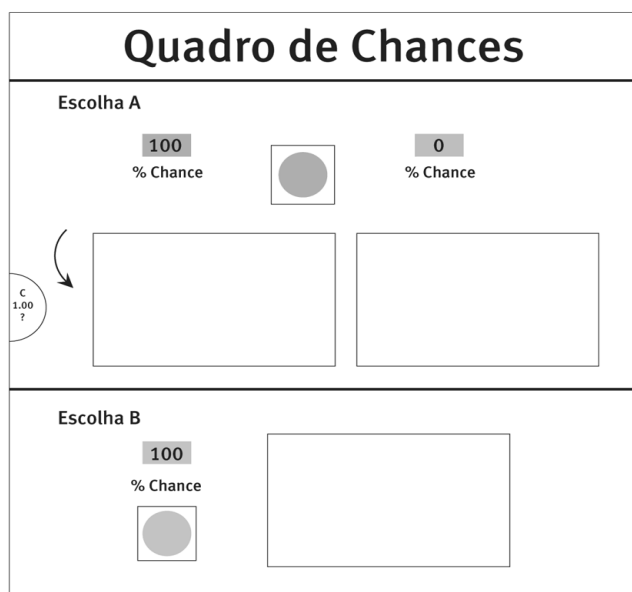


Fig. 1 - Representação do material utilizado para realizar *standard gamble*, o “Quadro de Chances”. O quadro é composto por um círculo que gira manualmente e as chances são sistematicamente variadas e demonstradas nas janelas superiores da “Resposta A”.

Os mesmos estados de saúde que o entrevistado irá ordenar na tarefa anterior serão utilizados para o *standard gamble*, agora representados por cartões coloridos. Os estados intermediários terão a cor verde e ficarão em um bolso no local da “Escolha B” do “Quadro de chances”, mostrando 100% de chance de permanecer naquele estado. A “Escolha A” será a “aposta” que demonstra do lado esquerdo a chance da intervenção funcionar e a pessoa recuperar a saúde plena e do lado direito as chances do desfecho oposto. Os cartões do melhor e pior desfecho serão apresentados nas cores rosa e azul, respectivamente. As chances serão apresentadas nas mesmas cores dos cartões, em rosa e azul, em números acima dos cartões e também na forma de “gráfico de torta”.

Após o término do SG, o entrevistador registrará o grau de dificuldade encontrado pelo entrevistado ao realizar o método.

Algoritmo para o SF-6D

A partir dos valores das preferências obtidos pelo *standard gamble*, modelos de regressão são construídos com os valores ajustados. Estes modelos geram um coeficiente de regressão para cada nível de cada um dos 6 domínios do SF-6D. Estes coeficientes representam o “peso” das preferências por aquele nível e são utilizados para construir a fórmula para cálculo dos índices de utilidade.

A modelagem também seguirá o mesmo protocolo da realizada pelo estudo britânico, considerando a possibilidade de modelos alternativos conforme os resultados encontrados. Os modelos serão feitos utilizando os dados de duas formas: considerando a avaliação de cada estado de saúde por cada entrevistado como uma observação independente (nível individual) e a média amostral para cada um dos 248 estados de saúde (nível agregado).

Várias tentativas serão efetuadas para que se encontre o modelo com melhor desempenho, mas de uma forma geral o modelo pode ser definido pela fórmula:

$$y_{ij} = g(\beta'x_{ij} + \theta r_{ij}) + \varepsilon_{ij} \text{ (modelo 1)}$$

onde $i = 1, 2, \dots, m$ representa o valor atribuído a cada estado de saúde e $j = 1, 2, \dots, n$ representa os indivíduos. A variável dependente, y_{ij} , é o valor ajustado do escore do *standard gamble* para o estado de saúde i avaliado pelo indivíduo j (SGADJ), x é um vetor de variáveis explicatórias dummies ($x_{\delta\lambda}$) para cada nível λ da dimensão δ do SF-6D. Por exemplo, x_{31} significa que a variável refere-se ao domínio 3 no nível 1. Para qualquer estado de saúde, $x_{\delta\lambda}$ será definido como:

$$\begin{aligned} x_{\delta\lambda} &= 1, \text{ se, para este estado, a dimensão } \delta \text{ está no nível } \lambda \\ x_{\delta\lambda} &= 0, \text{ se, para este estado, a dimensão } \delta \text{ não está no nível } \lambda \end{aligned}$$

No total haverá 25 variáveis, sendo o nível $\lambda = 1$ o valor basal para cada domínio. Considerando um modelo linear simples, o intercepto é representado pelo estado de saúde perfeita. O r é um vetor que considera as interações entre os níveis dos diferentes domínios; g é uma função especificando a forma funcional apropriada; ε_{ij} é um termo de erro cuja estrutura de autocorrelação e distribuição depende das suposições de cada modelo construído.

Este modelo é um modelo aditivo, mas que não impõe nenhuma restrição na relação entre os níveis de cada dimensão do SF-6D, como por exemplo, uma propriedade intervalar entre eles.

Seguindo a estratégia utilizada por Brazier et al. (7), os modelos serão avaliados considerando os seguintes critérios: a) inconsistência nos coeficientes: dada a estrutura ordinal dos níveis do SF-6D, onde o primeiro nível é o melhor e o último o pior em termos de disfunção, espera-se que os coeficientes, a partir do intercepto, aumentem em tamanho absoluto do segundo até o último nível e tenham o sinal negativo, pois representam perdas de função em relação à constante. Se isto não ocorre, é considerado uma inconsistência. Esta regra é válida para coeficientes com significância estatística; b) capacidade dos modelos para predizer todos os possíveis índices de utilidade para todos os estados de saúde. A habilidade preditiva será avaliada através: a) erro médio absoluto (MAE), b) proporção de predições fora da faixa entre 0.05 (%AE > 0.05) e 0.10 (%AE > 0.10), c) vieses de predição através de teste t , d) normalidade dos resíduos (teste de Jarque-Bera – JB), e) presença de autocorrelação nos erros de predição (estatística Ljung-Box- LB). Todas as análises serão realizadas nos programas *Statistical Package for Social Science* – SPSS (SPSS Inc. IBM company, Chicago) versão 16.0(12), R 2.9.1(Vienna)(13) e STATA 9.0 (Stata Corp. Texas)(14) .

Componente 2:

Este componente do estudo será feito concomitante ao Componente 1 e consiste na aplicação dos instrumentos SF-36 e WHOQOL-breve na população de estudo.

Descrição dos instrumentos

1) O SF-36 é um instrumento genérico cuja base conceitual para seu desenvolvimento é o de “qualidade de vida relacionada à saúde”. Os conceitos de saúde escolhidos para a construção da medida foram funcionamento físico, social, funcionamento no papel (*role functioning*), percepções da saúde geral, saúde mental, dor e vitalidade. Este construto está representado por 36 questões divididas em oito domínios: capacidade funcional, aspectos físicos, dor, estado geral de saúde, vitalidade, aspectos sociais, aspectos emocionais e saúde mental. A pontuação dos itens é feita através de uma escala de Likert. Todos os itens do SF-36 são usados para pontuar os oito domínios, exceto um, o item 2, que se refere a um auto-relato de transição de saúde. Cada item faz parte de somente um domínio. Após recalibrar dois itens e inverter o escore de nove itens, as respostas aos itens são somadas. Escores mais altos representam melhores estados de saúde. Um escore para *missing* é computado se itens de uma escala não foram respondidos. A escala dos escores vai de 0 a

100, 0 indicando o estado de saúde menos favorável, 100 o mais favorável e os escores entre eles representam o percentual do possível escore total alcançado. O SF-36 é um questionário que pode ser administrado por: auto-aplicação, administração por computador, por um entrevistador treinado pessoalmente ou por telefone e é adequado para pessoas acima dos 14 anos de idade. Pode ser administrado em 5 a 10 minutos com alto grau de aceitabilidade e qualidade dos dados (15). O SF-36 foi traduzido para a língua portuguesa e validado no Brasil por Ciconelli e cols. (16).

2) O WHOQOL-breve (17) é um instrumento genérico de qualidade de vida composto por 26 questões, sendo duas questões gerais de qualidade de vida e as demais 24 representam cada uma das 24 facetas que compõe o instrumento original, o WHOQOL-100 (18). As 24 questões estão divididas em 4 domínios: físico, psicológico, relações sociais e meio ambiente. Cada item é pontuado de 1 a 5 em uma escala do tipo Likert. Os escores são então transformados linearmente em uma escala entre 0 e 100, com 0 representando a pior qualidade de vida e 100 a melhor. Os valores de consistência interna (alfa de Cronbach) da versão brasileira do WHOQOL-breve relatados pelo estudo de validação foram de: 0,77 para todos os domínios, 0,91 para as 26 questões, 0,84 para o domínio físico, 0,79 para o domínio psicológico, 0,69 para o domínio relações sociais e 0,71 para o domínio meio-ambiente (1).

A análise estatística realizada com os dados obtidos da aplicação destes instrumentos será de natureza descritiva. As comparações de médias de acordo com variáveis sociodemográficas serão feitas por ANOVA. Para todos os testes o nível de significância será <0.05. Os dados serão analisados através dos programas SPSS (IBM Company, Chicago)(12), version 13.0 and Microsoft Office Excel 2003.

Protocolo de entrevistas

Uma equipe de entrevistadores composta por estudantes de graduação de cursos da área da saúde receberá treinamento intensivo para aplicação da técnica do *standard gamble* e a seguir a rotina de aplicação do restante do protocolo. As entrevistas serão realizadas no domicílio dos participantes e a apresentação da equipe ocorrerá de uma carta onde serão descritos o propósito do estudo, o tipo de tarefa que o entrevistado irá desenvolver e o tempo despendido na entrevista. O protocolo consiste dos seguintes instrumentos, que serão aplicados estritamente nesta ordem: 1. identificação do entrevistado; 2. aplicação do instrumento SF-6D para que o participante tenha familiaridade com a descrição de saúde que irá ser apresentada nos cartões contendo os estados de saúde; 3. ordenamento dos 8 estados de saúde de acordo com a preferência do indivíduo, aplicação do *standard gamble* e verificação do grau de dificuldade encontrado nesta tarefa; 4. aplicação dos instrumentos de medida de qualidade de vida SF-36 e WHOQOL-breve ; 5. questionário sociodemográfico.

Antes da entrevista, todos os participantes assinaram Termo de Consentimento Informado aprovado pelo Comitê de Ética da universidade.

Seleção da amostra

A seleção da amostra será feita por um plano amostral complexo, por conglomerados. Estes serão representados pelos setores censitários da cidade de Porto Alegre, segundo divisão feita pelo IBGE (19). Tanto a seleção de setores censitários como dos domicílios a serem visitados em cada setor serão selecionados randomicamente.

O número de domicílios a ser visitado será calculado pela seguinte fórmula (20):

$$No.domiclios = \frac{no.de _ indivíduos _ por _ sexo _ e _ faixa _ etaria}{no._ de _ pessoas _ por _ domicilio * prop.de _ indiv.na _ pop._ por _ sexo _ e _ fai _ etaria}$$

A proporção de indivíduos de acordo com sexo e faixa etária na população e o número de pessoas por domicílio serão obtidos através das informações do IBGE (19).

A fim de calcular quantos domicílios deverão ser visitados por setor e conseqüentemente quantos setores serão necessários, utilizaremos uma abordagem descrita por Barros e cols.(20) que

considera o número possível de entrevistas que podem ser realizadas em um dia de trabalho por um grupo de entrevistadores. Em cada setor, os quarteirões serão numerados e um deles será sorteado para as visitas. No quarteirão sorteado, a abordagem aos domicílios será feita da seguinte forma: a partir da esquina sorteada, o entrevistador deve andar para a esquerda de quem está de frente para a casa visitando os domicílios a cada determinado número de casas ou apartamentos. Este número será estimado de forma aproximada dividindo-se o total de domicílios no setor (informação obtida através do IBGE) pela quantidade de quarteirões e após pelo número de casas que devem ser incluídas. Por exemplo, um setor censitário tem 300 domicílios e 8 quadras. Se o número de domicílios a serem visitados é de 4, então o entrevistador devera visitar uma residência a cada 9 ($300/8 = 37,5$ e $37,5/4 = 9,3$). Exceção a esta regra ocorrerá caso o setor tenha quadras ocupadas por praças, parques ou terrenos baldios. Caso o domicílio sorteado seja ponto comercial, o entrevistador deve abordar a próxima residência particular.

Todos os indivíduos moradores serão convidados a participar desde que preencham os seguintes critérios de inclusão: ter entre 20 e 64 anos de idade, ser alfabetizado e não ser portador de nenhuma doença física ou mental que impedisse o auto-preenchimento dos instrumentos.

5 CRONOGRAMA

Ano 2006

Atividades	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Definição do projeto – encaminhamento CEPs	■	■										
Revisão de Literatura		■	■	■	■							
Intercâmbio internacional – visita pesquisador do grupo à Universidade de Scheffield Health Economics – UK						■	■	■				
Contratação e treinamento equipe de coleta de dados									■	■	■	■

Ano 2007

Atividades	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Confeccção do material para entrevistas e seleção de equipe de entrevistadores	■	■	■	■								
Treinamento da equipe de coleta					■							
Aplicação dos questionários – coleta de dados						■	■	■	■	■	■	■

6 ORÇAMENTO

ITEM DE DISPÊNDIO	DESCRIÇÃO	VALOR TOTAL (R\$)
Equipamentos e Material permanente	Impressora Laser Lexmark E230 (monocromática, 1.200 Image Quality, 600x600 dpi, 18 ppm, ciclo de vida mensal de 10.000 páginas, 110 V, 1 ano de garantia Lexmark balcão) - valor unitário R\$ 648,70	648,70
Diárias	20 (vinte) Diárias no exterior (intercambio aluna de Pos-Graduação na Universidade de Sheffield - = R\$ 5.670,00 20 (vinte) diárias no país (intercambio de 2 pesquisadores da Universidade de Sheffield - = R\$ 2.969,00 12 (doze) diárias no país	8.639,00
Passagens	1 (uma) passagem Porto Alegre - Reino Unido - Porto Alegre (intercambio aluna do PPG com Universidade de Sheffield - R\$ 3.460,51 (unitario US 1281,67) 2 (duas) passagens Reino Unido - Porto Alegre - Reino Unido (intercambio Universidade de Sheffield - R\$ 6.921,02 (US 1281,67)	10.381,53
Material de consumo (Total)	Material de consumo geral - folhas A4, reprografias, revelações, tonner impressora, busca artigos BIREME. Estimativa média mensal R\$ 300,00 - total projeto R\$ 7.200,00	7.200,00
Equipamentos e Material permanente	Notebook - LAPTOP dell LatitudeTM 110L (BR6606) Processador Intel® Celeron® M 350 (1,30GHz, 1 MB L2 Cache) Microsoft Windows XP Home SP2 em Português Bateria Lithium Íon 256MB de Memória PC2700 DDR SDRAM valor unitário R\$ 3.999,00	3.999,00
Pessoal	Contratação entrevistadores para aplicação de questionários (busca ativa) e locomoção	9.000,00
Total geral		39.868,23

7 QUESTÕES ÉTICAS

A participação dos indivíduos no presente estudo consiste em uma entrevista com aplicação de instrumentos auto-aplicáveis e também aplicados pelo entrevistador. Tal procedimento implicará no envolvimento dos participantes na pesquisa pelo período de aproximadamente 50 minutos, informação que será fornecida às pessoas antes de concordarem em participar do estudo.

Todos os participantes deverão assinar o Termo de Consentimento antes da entrada no estudo.

Este projeto de pesquisa será encaminhado para análise pelo Comitê de Ética do Hospital de Clínicas de Porto Alegre.

Referências bibliográficas

- (1) Fleck MP, Louzada S, Xavier M, Chachamovich E, Vieira G, Santos L, et al. [Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref]. *Rev Saude Publica* 2000 Apr;34(2):178-83.
- (2) Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993 Apr 15;118(8):622-9.
- (3) Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care* 2000 Sep;38(9 Suppl):II138-II150.
- (4) Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996 Oct 16;276(15):1253-8.
- (5) Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996 Oct 9;276(14):1172-7.
- (6) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996 Aug 3;313(7052):275-83.
- (7) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002 Mar;21(2):271-92.
- (8) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 Jun;30(6):473-83.
- (9) Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Revista Brasileira de Reumatologia* 1999;39(3):143-50.
- (10) von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. First ed. Princeton: Princeton University Press; 1944.
- (11) Furlong W, Feeny D, Torrance G, Barr R, Horsman J. *Guide to Design and Development of Health-State Utility Instrumentation*. [Paper #90-9]. 1990. Canada, Centre for Health Economic and Policy Analysis. McMaster University. Working Paper Series. Ref Type: Serial (Book, Monograph).
- (12) *Statistical Package for Social Sciences*. IBM Company 2010 March 6 Available from: URL: <http://www.spss.com/software/>.
- (13) R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing 2010 March 6(2009) Available from: URL: www.R-project.org.
- (14) *Data analysis and statistical software*. StataCorp LP 2010 March 6 Available from: URL: www.stata.com.
- (15) Ware JE, Jr. SF-36 health survey update. *Spine (Phila Pa 1976)* 2000 Dec 15;25(24):3130-9.
- (16) Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras Reumatol* 1999;39(3):143-50.

- (17) Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. *Psychol Med* 1998 May;28(3):551-8.
- (18) The WHOQOL group. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Social Science and Medicine* 1995;41(10):1403-9.
- (19) Censo Populacional 2000. Instituto Brasileiro de Geografia e Estatística 2010 March 6 Available from: URL: www.ibge.gov.br.
- (20) Barros FC, Victora CG. *Epidemiologia da Saúde Infantil*. First ed. São Paulo: HUCITEC-UNICEF; 1991.

ANEXO 2

TERMO DE CONSENTIMENTO INFORMADO

Projeto Colaborativo Interinstitucional de Estudos de Custo-Efetividade em Saúde
Medidas de Qualidade de Vida e Utilidade em uma amostra da população brasileira

Termo de Consentimento

O Sr (a). está sendo convidado a participar de uma pesquisa que avaliará a qualidade de vida de uma amostra da população brasileira. Solicitamos ao Sr (a). que leia este documento e, se tiver qualquer dúvida, estaremos à disposição para esclarecê-lo antes do Sr (a). concordar ou não em participar.

O estudo está sendo conduzido por pesquisadores que são alunos e professores dos cursos de pós-graduação em Epidemiologia e Psiquiatria da Universidade Federal do Rio Grande do Sul.

Qualidade de vida significa o quanto a pessoa está satisfeita com a saúde do seu corpo, com seu estado emocional, com a realização de suas tarefas e com suas relações sociais. Ela pode ser avaliada através de questionários que contêm várias perguntas sobre estes diferentes aspectos da vida das pessoas: sua saúde física e emocional, sua relação com amigos e familiares, seu meio-ambiente.

Desta forma, gostaríamos de contar com a sua colaboração para responder a três questionários que avaliarão sua qualidade de vida. O tempo dispendido será de aproximadamente 50 minutos. O Sr (a). terá a liberdade de recusar-se a responder alguma pergunta, se assim quiser. Todas as instruções necessárias para as respostas aos instrumentos serão fornecidas pelo entrevistador.

A colaboração no estudo não implica em nenhum risco para o Sr (a), uma vez que só será necessário responder aos questionários. A colaboração nesta pesquisa é voluntária. Portanto, o Sr. (a) tem toda a liberdade de recusar-se a participar. Todas as informações sobre a sua pessoa são confidenciais, e seu nome não aparecerá na divulgação dos resultados do trabalho.

Para contatos e dúvidas, nome e telefones da coordenadora do projeto:

Dra. Luciane N. Cruz – 9221-9890 ou 2101-8843

O Sr (a) pode perguntar agora qualquer dúvida que tenha tido.

Será fornecida ao Sr(a). uma cópia deste documento.

Consentimento :

“ Eu li todas as informações acima. Fiz as perguntas necessárias e obtive respostas. Eu concordo em participar do estudo”

Assinatura : _____ Data : _____

Assinatura do pesquisador : _____ Data : _____

ANEXO 3

PROTOCOLO DO ESTUDO

--	--	--	--

Hora de Início SF6D:

		:		
--	--	---	--	--

Questionário auto-aplicável

SF-6D

As perguntas nas próximas páginas questionam sobre diferentes aspectos da sua saúde. Há seis grupos com diferentes condições, cada grupo relaciona-se a um determinado aspecto de saúde. Por favor, marque uma condição em cada grupo a fim de mostrar qual delas melhor descreve sua própria condição de saúde nas últimas quatro semanas.

Por favor, marque uma condição em cada grupo a fim de mostrar qual delas melhor descreve sua própria saúde.

Marque apenas uma resposta

Nível - Capacidade funcional

- 1 - Sua saúde não dificulta que você faça atividades vigorosas, que exigem muito esforço, tais como correr, levantar objetos pesados, participar em esportes árduos
- 2 - Sua saúde dificulta um pouco que você faça atividades vigorosas, que exigem muito esforço, tais como correr, levantar objetos pesados, participar em esportes árduos
- 3 - Sua saúde dificulta um pouco que você faça atividades moderadas, tais como mover uma mesa, passar aspirador de pó, jogar bola, varrer a casa.
- 4 - Sua saúde dificulta muito que você faça atividades moderadas, tais como mover uma mesa, passar aspirador de pó, jogar bola, varrer a casa.
- 5 - Sua saúde dificulta um pouco para você tomar banho ou vestir-se.
- 6 - Sua saúde dificulta MUITO para você tomar banho ou vestir-se

Nível - Aspectos físicos e emocionais

- 1- Você não teve problemas com o seu trabalho ou alguma outra atividade diária regular como consequência de sua saúde física ou algum problema emocional (como sentir-se deprimido ou ansioso)
- 2 - Você esteve limitado no seu tipo de trabalho ou em outras atividades como consequência de sua saúde física
- 3 - Você realizou menos tarefas do que você gostaria como consequência de algum problema emocional
- 4 - Você esteve limitado no seu tipo de trabalho ou em outras atividades como consequência de sua saúde física e realizou menos tarefas do que você gostaria como consequência de algum problema emocional

--	--	--	--

Nível - Aspectos sociais

- 1 - Sua saúde física ou problemas emocionais não interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) em nenhuma parte do tempo
- 2 - Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) em uma pequena parte do tempo
- 3 - Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) em alguma parte do tempo
- 4 - Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) na maior parte do tempo
- 5 - Sua saúde física ou problemas emocionais interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) todo o tempo

Nível - Dor

- 1 - Você não teve nenhuma dor no corpo
- 2 - Você teve dor, mas a dor não interferiu de maneira alguma com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
- 3 - Você teve dor, e a dor interferiu um pouco com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
- 4 - Você teve dor, e a dor interferiu moderadamente com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
- 5 - Você teve dor, e a dor interferiu bastante com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)
- 6 - Você teve dor, e a dor interferiu extremamente com o seu trabalho normal (incluindo tanto o trabalho fora de casa e dentro de casa)

Nível - Saúde mental

- 1 - Você não tem se sentido uma pessoa muito nervosa ou desanimada e abatida nunca
- 2 - Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida em uma pequena parte do tempo
- 3 - Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida em alguma parte do tempo
- 4 - Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida na maior parte do tempo
- 5 - Você tem se sentido uma pessoa muito nervosa ou desanimada e abatida todo o tempo

--	--	--	--

Nível - Vitalidade

- 1 - Você tem se sentido com muita energia todo o tempo
- 2 - Você tem se sentido com muita energia a maior parte do tempo
- 3 - Você tem se sentido com muita energia em alguma parte do tempo
- 4 - Você tem se sentido com muita energia em uma pequena parte do tempo
- 5 - Você não tem se sentido com muita energia nunca

Hora de Término SF6D:

		:		
--	--	---	--	--

Tempo total:

--	--

--	--	--	--

Termômetro + Quadro de chances

Hora de Início Termômetro: :

Termômetro

Cartão com o estado de saúde:	Classificação:
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/>

Hora de Término Termômetro: :

Tempo Total:

Grade de Respostas - Quadro de Chances

Hora de Início Quadro de Chances: :

Número do Cartão:

C1.1: (Rosa 100% / Azul 0%)

- A **Ir para C1.2**
- B **Ir para C1.13**
- Não consigo decidir **Ir para C2**

C1.3: (Cobertura 1)

- A **Ir para C1.14**
- B **Ir para C2**
- Não consigo decidir **Ir para C2**

C1.5: (Cobertura 2)

- A **Ir para C2**
- B **Ir para C2**
- Não consigo decidir **Ir para C2**

C1.2: (Rosa 10% / Azul 90%)

- A **Ir para C1.3**
- B **Ir para C1.4**
- Não consigo decidir **Ir para C2**

C1.4: (Rosa 90% / Azul 10%)

- A **Ir para C1.6**
- B **Ir para C1.5**
- Não consigo decidir **Ir para C2**

C1.6: (Rosa 20% / Azul 80%)

- A **Ir para C2**
- B **Ir para C1.7**
- Não consigo decidir **Ir para C2**

--	--	--	--

C1.7: (Rosa 80% / Azul 20%)

- A **Ir para C1.8**
 B **Ir para C2**
 Não consigo decidir **Ir para C2**

C1.9: (Rosa 70% / Azul 30%)

- A **Ir para C1.10**
 B **Ir para C2**
 Não consigo decidir **Ir para C2**

C1.11: (Rosa 60% / Azul 40%)

- A **Ir para C1.12**
 B **Ir para C2**
 Não consigo decidir **Ir para C2**

C1.8: (Rosa 30% / Azul 70%)

- A **Ir para C2**
 B **Ir para C1.9**
 Não consigo decidir **Ir para C2**

C1.10: (Rosa 40% / Azul 60%)

- A **Ir para C2**
 B **Ir para C1.11**
 Não consigo decidir **Ir para C2**

C1.12: (Rosa 50% / Azul 50%)

- A **Ir para C2**
 B **Ir para C2**
 Não consigo decidir **Ir para C2**

Utility: ,

C1.13: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C1.14: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Número do Cartão:

C2.1: (Rosa 100% / Azul 0%)

- A **Ir para C2.2**
 B **Ir para C2.13**
 Não consigo decidir **Ir para C3**

C2.3: (Cobertura 1)

- A **Ir para C2.14**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

C2.5: (Cobertura 2)

- A **Ir para C3**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

C2.2: (Rosa 10% / Azul 90%)

- A **Ir para C2.3**
 B **Ir para C2.4** **Ir para C3**
 Não consigo decidir

C2.4: (Rosa 90% / Azul 10%)

- A **Ir para C2.6**
 B **Ir para C2.5**
 Não consigo decidir **Ir para C3**

C2.6: (Rosa 20% / Azul 80%)

- A **Ir para C3**
 B **Ir para C2.7**
 Não consigo decidir **Ir para C3**

--	--	--	--

C2.7: (Rosa 80% / Azul 20%)

- A **Ir para C2.8**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

C2.9: (Rosa 70% / Azul 30%)

- A **Ir para C2.10**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

C2.11: (Rosa 60% / Azul 40%)

- A **Ir para C2.12**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

C2.8: (Rosa 30% / Azul 70%)

- A **Ir para C3**
 B **Ir para C2.9**
 Não consigo decidir **Ir para C3**

C2.10: (Rosa 40% / Azul 60%)

- A **Ir para C3**
 B **Ir para C2.11**
 Não consigo decidir **Ir para C3**

C2.12: (Rosa 50% / Azul 50%)

- A **Ir para C3**
 B **Ir para C3**
 Não consigo decidir **Ir para C3**

Utility: ,

C2.13: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C2.14: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Número do Cartão:

N° do Quest

C3.1: (Rosa 100% / Azul 0%)

- A **Ir para C3.2**
 B **Ir para C3.13**
 Não consigo decidir **Ir para C4**

C3.2: (Rosa 10% / Azul 90%)

- A **Ir para C3.3**
 B **Ir para C3.4**
 Não consigo decidir **Ir para C4**

C3.3: (Cobertura 1)

- A **Ir para C3.14**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

C3.4: (Rosa 90% / Azul 10%)

- A **Ir para C3.6**
 B **Ir para C3.5**
 Não consigo decidir **Ir para C4**

C3.5: (Cobertura 2)

- A **Ir para C4**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

C3.6: (Rosa 20% / Azul 80%)

- A **Ir para C4**
 B **Ir para C3.7**
 Não consigo decidir **Ir para C4**

C3.7: (Rosa 80% / Azul 20%)

- A **Ir para C3.8**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

C3.8: (Rosa 30% / Azul 70%)

- A **Ir para C4**
 B **Ir para C3.9**
 Não consigo decidir **Ir para C4**

C3.9: (Rosa 70% / Azul 30%)

- A **Ir para C3.10**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

C3.10: (Rosa 40% / Azul 60%)

- A **Ir para C4**
 B **Ir para C3.11**
 Não consigo decidir **Ir para C4**

C3.11: (Rosa 60% / Azul 40%)

- A **Ir para C3.12**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

C3.12: (Rosa 50% / Azul 50%)

- A **Ir para C4**
 B **Ir para C4**
 Não consigo decidir **Ir para C4**

Utility: ,

C3.13: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C3.14: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Número do Cartão:

Nº do Quest

C4.1: (Rosa 100% / Azul 0%)

- A **Ir para C4.2**
 B **Ir para C4.13**
 Não consigo decidir **Ir para C5**

C4.2: (Rosa 10% / Azul 90%)

- A **Ir para C4.3**
 B **Ir para C4.4**
 Não consigo decidir **Ir para C5**

C4.3: (Cobertura 1)

- A **Ir para C4.14**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

C4.4: (Rosa 90% / Azul 10%)

- A **Ir para C4.6**
 B **Ir para C4.5**
 Não consigo decidir **Ir para C5**

C4.5: (Cobertura 2)

- A **Ir para C5**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

C4.6: (Rosa 20% / Azul 80%)

- A **Ir para C5**
 B **Ir para C4.7**
 Não consigo decidir **Ir para C5**

C4.7: (Rosa 80% / Azul 20%)

- A **Ir para C4.8**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

C4.8: (Rosa 30% / Azul 70%)

- A **Ir para C5**
 B **Ir para C4.9**
 Não consigo decidir **Ir para C5**

C4.9: (Rosa 70% / Azul 30%)

- A **Ir para C4.10**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

C4.10: (Rosa 40% / Azul 60%)

- A **Ir para C5**
 B **Ir para C4.11**
 Não consigo decidir **Ir para C5**

C4.11: (Rosa 60% / Azul 40%)

- A **Ir para C4.12**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

C4.12: (Rosa 50% / Azul 50%)

- A **Ir para C5**
 B **Ir para C5**
 Não consigo decidir **Ir para C5**

Utility: ,

C4.13: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C4.14: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Número do Cartão:

Nº do Quest

C5.1: (Rosa 100% / Azul 0%)

- A Ir para C5.2
 B Ir para C5.13
 Não consigo decidir Ir para C5.15

C5.2: (Rosa 10% / Azul 90%)

- A Ir para C5.3
 B Ir para C5.4
 Não consigo decidir Ir para C5.15

C5.3: (Cobertura 1)

- A Ir para C5.14
 B Ir para C5.15
 Não consigo decidir Ir para C5.15

C5.4: (Rosa 90% / Azul 10%)

- A Ir para C5.6
 B Ir para C5.5
 Não consigo decidir Ir para C5.15

C5.5: (Cobertura 2)

- A Ir para C5.15
 B Ir para C5.15
 Não consigo decidir Ir para C5.15

C5.6: (Rosa 20% / Azul 80%)

- A Ir para C5.15
 B Ir para C5.7
 Não consigo decidir Ir para C5.15

C5.7: (Rosa 80% / Azul 20%)

- A Ir para C5.8
 B Ir para C5.15
 Não consigo decidir Ir para C.15

C5.8: (Rosa 30% / Azul 70%)

- A Ir para C5.15
 B Ir para C5.9
 Não consigo decidir Ir para C5.15

C5.9: (Rosa 70% / Azul 30%)

- A Ir para C5.10
 B Ir para C5.15
 Não consigo decidir Ir para C5.15

C5.10: (Rosa 40% / Azul 60%)

- A Ir para C5.15
 B Ir para C5.11
 Não consigo decidir Ir para C5.15

C5.11: (Rosa 60% / Azul 40%)

- A Ir para C5.12
 B Ir para C5.15
 Não consigo decidir Ir para C5.15

C5.12: (Rosa 50% / Azul 50%)

- A Ir para C5.15
 B Ir para C5.15
 Não consigo decidir Ir para C5.15

Utility: ,

C5.13: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C5.14: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

--	--	--	--

C5.15: Visto do entrevistador

- 1- O entrevistado classificou o CARTÃO M acima do CARTÃO EP
- 2- O entrevistado classificou o CARTÃO EP acima do CARTÃO M
- 3- O entrevistado classificou o CARTÃO EP no mesmo nível que o CARTÃO M

1-IR PARA C6a

2-IR PARA C6b

3-IR PARA D

--	--	--	--

C6a:**Cartão azul no bolso direito da opção A e Cartão M no bolso da opção B****C6a.1:**

- Sim **Ir para C6a.3**
- Não **Ir para C6a.2**
- Não sei **Ir para D**

C6a.3: (Rosa 100%/Azul 0%)

- A **Ir para C6a.4**
- B **Ir para C6a.15**
- Não consigo decidir **Ir para D**

C6a.5: (Cobertura 1)

- A **Ir para C6a.16**
- B **Ir para D**
- Não consigo decidir **Ir para D**

C6a.7: (Cobertura 2)

- A **Ir para D**
- B **Ir para D**
- Não consigo decidir **Ir para D**

C6a.9: (Rosa 80% / Azul 20%)

- A **Ir para C6a.10**
- B **Ir para D**
- Não consigo decidir **Ir para D**

C6a.11: (Rosa 70% / Azul 30%)

- A **Ir para C6a.12**
- B **Ir para D**
- Não consigo decidir **Ir para D**

C6a.13: (Rosa 60% / Azul 40%)

- A **Ir para C6a.14**
- B **Ir para D**
- Não consigo decidir **Ir para D**

C6a.2:

- Sim **Ir para C6b**
- Não **Ir para D**
- Não sei **Ir para D**

C6a.4: (Rosa 10% / Azul 90%)

- A **Ir para C6a.5**
- B **Ir para C6a.6**
- Não consigo decidir **Ir para D**

C6a.6: (Rosa 90% / Azul 10%)

- A **Ir para C6a.8**
- B **Ir para C6a.7**
- Não consigo decidir **Ir para D**

C6a.8: (Rosa 20% / Azul 80%)

- A **Ir para D**
- B **Ir para C6a.9**
- Não consigo decidir **Ir para D**

C6a.10: (Rosa 30% / Azul 70%)

- A **Ir para D**
- B **Ir para C6a.11**
- Não consigo decidir **Ir para D**

C6a.12: (Rosa 40% / Azul 60%)

- A **Ir para D**
- B **Ir para C6a.13**
- Não consigo decidir **Ir para D**

C6a.14: (Rosa 50% / Azul 50%)

- A **Ir para D**
- B **Ir para D**
- Não consigo decidir **Ir para D**

Utility: ,

--	--	--	--

C6a.15: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C6a.16: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Hora de Término Quadro de Chances: :

Tempo Total Quadro de Chances:

C7: Como você considerou o exercício do quadro de chances? (LEIA AS ALTERNATIVAS)

Muito Fácil Fácil Um pouco difícil Difícil Muito difícil

AGORA IR PARA D

--	--	--	--

C6b:**Cartão M no bolso direito da opção A e Cartão azul no bolso da opção B****C6b.1:**

- Sim **Ir para C6b.3**
 Não **Ir para C6b.2**
 Não sei **Ir para D**

C6b.2:

- Sim **Ir para C6a.3**
 Não **Ir para D**
 Não sei **Ir para D**

C6b.3: (Rosa 100%/Azul 0%)

- A **Ir para C6b.4**
 B **Ir para C6b.15**
 Não consigo decidir **Ir para D**

C6b.4: (Rosa 10% / Azul 90%)

- A **Ir para C6b.5**
 B **Ir para C6b.6**
 Não consigo decidir **Ir para D**

C6b.5: (Cobertura 1)

- A **Ir para C6b.16**
 B **Ir para D**
 Não consigo decidir **Ir para D**

C6b.6: (Rosa 90% / Azul 10%)

- A **Ir para C6b.8**
 B **Ir para C6b.7**
 Não consigo decidir **Ir para D**

C6b.7: (Cobertura 2)

- A **Ir para D**
 B **Ir para D**
 Não consigo decidir **Ir para D**

C6b.8: (Rosa 20% / Azul 80%)

- A **Ir para D**
 B **Ir para C6b.9**
 Não consigo decidir **Ir para D**

C6b.9: (Rosa 80% / Azul 20%)

- A **Ir para C6b.10**
 B **Ir para D**
 Não consigo decidir **Ir para D**

C6b.10: (Rosa 30% / Azul 70%)

- A **Ir para D**
 B **Ir para C6b.11**
 Não consigo decidir **Ir para D**

C6b.11: (Rosa 70% / Azul 30%)

- A **Ir para C6b.12**
 B **Ir para D**
 Não consigo decidir **Ir para D**

C6b.12: (Rosa 40% / Azul 60%)

- A **Ir para D**
 B **Ir para C6b.13**
 Não consigo decidir **Ir para D**

C6b.13: (Rosa 60% / Azul 40%)

- A **Ir para C6a.14**
 B **Ir para D**
 Não consigo decidir **Ir para D**

C6b.14: (Rosa 50% / Azul 50%)

- A **Ir para D**
 B **Ir para D**
 Não consigo decidir **Ir para D**

Utility: ,

--	--	--	--

C6b.15: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão verde ao invés de 100% de chance de estar no estado de saúde descrito no cartão rosa?

C6b.16: Por que você escolheu 100% de chance de estar em estado de saúde indicado no cartão azul ao invés de 100% de chance de estar no estado de saúde descrito no cartão verde?

Hora de Término Quadro de Chances: :

Tempo Total Quadro de Chances:

C7: Como você considerou o exercício do quadro de chances? (LEIA AS ALTERNATIVAS)

Muito Fácil Fácil Um pouco difícil Difícil Muito difícil

AGORA IR PARA D

--	--	--	--

D. Aplicação dos questionários SF-36 e WHOQOL-abreviado

Hora de Início SF-36: :

Versão brasileira do questionário de Qualidade de Vida SF-36

Esta pesquisa questiona você sobre sua saúde. Estas informações nos manterão informados de como você se sente e quão bem você é capaz de fazer suas atividades de vida diária. Responda cada questão marcando a resposta como indicado. Caso você esteja inseguro em como responder, por favor tente responder o melhor que puder.

Marque com X uma alternativa

1-Em geral, você diria que a sua saúde é:

Excelente	Muito Boa	Boa	Ruim	Muito Ruim
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

2- Comparada a um ano atrás, como você classificaria sua saúde em geral, agora ?

Muito melhor	Um pouco melhor	Quase a mesma	Um pouco pior	Muito pior
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

--	--	--	--

3- Os seguintes itens são sobre atividades que você poderia fazer atualmente durante um dia comum. **Devido a sua saúde**, você tem dificuldade para fazer essas atividades? Neste caso, quanto?

ATIVIDADES	SIM. DIFICULTA MUITO	SIM. DIFICULTA UM POUCO	NÃO. NÃO DIFICULTA DE MODO ALGUM
a) Atividades vigorosas , que exigem muito esforço, tais como correr, levantar objetos pesados, participar em esportes árduos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
b) Atividades moderadas , tais como mover uma mesa, passar aspirador de pó, jogar bola, varrer a casa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
c) Levantar ou carregar mantimentos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
d) Subir vários lances de escada	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
e) Subir um lance de escada	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
f) Curvar-se, ajoelhar-se ou dobrar-se	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
g) Andar mais de 1 quilômetro	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
h) Andar vários quarteirões	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
i) Andar um quarteirão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
j) Tomar banho ou vestir-se	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3

4- Durante as **últimas 4 semanas**, você teve algum dos seguintes problemas com o seu trabalho ou com alguma atividade diária regular, **como consequência de sua saúde física** ?

	Sim	Não
a) Você diminuiu a quantidade de tempo que se dedicava ao seu trabalho ou a outras atividades ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2
b) Realizou menos tarefas do que você gostaria ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2
c) Esteve limitado no seu tipo de trabalho ou em outras atividades ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2
d) Teve dificuldade de fazer seu trabalho ou outras atividades (por exemplo: necessitou de um esforço extra)?	<input type="checkbox"/> 1	<input type="checkbox"/> 2

--	--	--	--

5- Durante as últimas 4 semanas, você teve alguns dos seguintes problemas com o seu trabalho ou outra atividade regular diária, como consequência de algum problema emocional (como sentir-se deprimido ou ansioso) ?

	Sim	Não
a) Você diminui a quantidade de tempo que se dedicava ao seu trabalho ou a outras atividades ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2
b) Realizou menos tarefas do que você gostaria ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2
c) Não trabalhou ou fez qualquer das atividades com tanto cuidado como geralmente faz ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2

6- Durante as últimas **4 semanas**, de que maneira sua saúde física ou problemas emocionais interferiram nas suas atividades sociais normais , em relação a família , vizinhos, amigos ou em grupo ?

De forma nenhuma	Ligeiramente	Moderadamente	Bastante	Extremamente
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

7- Quanta dor **no corpo** você teve durante as últimas **4 semanas** ?

Nenhuma	Muito leve	Leve	Moderada	Grave	Muito grave
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

8- Durante as **últimas 4 semanas** , quanto a dor interferiu com o seu trabalho normal (incluindo tanto o trabalho, fora e dentro de casa) ?

De maneira nenhuma	Um pouco	Moderadamente	Bastante	Extremamente
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

--	--	--	--

9- Estas questões são sobre como você se sente e como tudo tem acontecido com você durante as **últimas 4 semanas**. Para cada questão, por favor dê uma resposta que mais se aproxime da maneira como você se sente. Em relação as **últimas 4 semanas**.

	Todo o tempo	A maior parte do tempo	Uma boa parte do tempo	Alguma parte do tempo	Uma pequena parte do tempo	Nunca
a) Quanto tempo você tem se sentido cheio de vigor, cheio de vontade, cheio de força ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
b) Quanto tempo você tem se sentido uma pessoa muito nervosa ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
c) Quanto tempo você tem se sentido tão deprimido que nada pode animá-lo ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
d) Quanto tempo você tem se sentido calmo ou tranqüilo ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
e) Quanto tempo você tem se sentido com muita energia ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
f) Quanto tempo você tem se sentido desanimado e abatido ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
g) Quanto tempo você tem se sentido esgotado ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
h) Quanto tempo você tem se sentido uma pessoa feliz ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
i) Quanto tempo você tem se sentido cansado ?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

10- Durante as **últimas 4 semanas** , quanto do seu tempo a sua **saúde física ou problemas emocionais** interferiram com as suas atividades sociais (como visitar amigos, parentes, etc) ?

Todo tempo	A maior parte do tempo	Alguma parte do tempo	Uma pequena parte do tempo	Nenhuma parte do tempo
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

--	--	--	--

11- O quanto **verdadeiro** ou **falso** é cada uma das afirmações para você ?

	Definitiva- mente verdadeiro	A maioria das vezes verdadeiro	Não Sei	A maioria das vezes falsa	Definitiva- mente Falsa
a) Eu costumo adoecer um pouco mais facilmente que as outras pessoas.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b) Eu sou tão saudável quanto qualquer pessoa que eu conheço.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c) Eu acho que a minha saúde vai piorar.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d) Minha saúde é excelente.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Hora de Término SF-36: :

Tempo Total SF-36:

--	--	--	--

Hora de Início WHOQOL:

--	--

 :

--	--

WHOQOL - Abreviado

Instruções

Este questionário é sobre como você se sente a respeito de sua qualidade de vida, saúde e outras áreas de sua vida. Por favor, responda a todas as questões . Se você não tem certeza sobre que resposta dar em uma questão, por favor, escolha entre as alternativas a que lhe parece mais apropriada. Esta, muitas vezes, poderá ser sua primeira escolha.

Por favor, tenha em mente seus valores, aspirações, prazeres e preocupações. Nós estamos perguntando o que você acha de sua vida, tomando como referência as duas últimas semanas .

--	--	--	--

Por favor, leia cada questão, veja o que você acha e marque o número que lhe parece a melhor resposta.

	Muito Ruim	Ruim	Nem Ruim Nem Boa	Boa	Muito Boa
1- Como você avaliaria sua qualidade de vida?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

	Muito Insatisfeito	Insatisfeito	Nem Satisfeito Nem Insatisfeito	Satisfeito	Muito Satisfeito
2- Quão satisfeito(a) você está com a sua saúde?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

As questões seguintes são sobre o quanto você tem sentido algumas coisas nas últimas duas semanas.

	Nada	Muito Pouco	Mais ou Menos	Bastante	Extremamente
3- Em que medida você acha que sua dor (física) impede você de fazer o que você precisa?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
4- O quanto você precisa de algum tratamento médico para levar sua vida diária?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
5- O quanto você aproveita a vida?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
6- Em que medida você acha que a sua vida tem sentido?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
7- O quanto você consegue se concentrar?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
8- Quão seguro(a) você se sente em sua vida diária?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
9- Quão saudável é o seu ambiente físico (clima, barulho, poluição, atrativos)?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

--	--	--	--

As questões seguintes perguntam sobre quão completamente você tem sentido ou é capaz de fazer certas coisas nestas últimas duas semanas.

	Nada	Muito Pouco	Médio	Muito	Completa mente
10- Você tem energia suficiente para seu dia-a-dia?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
11- Você é capaz de aceitar sua aparência física?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
12- Você tem dinheiro suficiente para satisfazer suas necessidades?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
13- Quão disponíveis para você estão as informações que precisa no seu dia-a-dia?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
14- Em que medida você tem oportunidades de atividade de lazer?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

As questões seguintes perguntam sobre quão bem ou satisfeito você se sentiu a respeito de vários aspectos de sua vida nas últimas duas semanas.

	Muito Ruim	Ruim	Nem Ruim Nem Boa	Boa	Muito Boa
15- Quão bem você é capaz de se locomover?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

--	--	--	--

	Muito Insatisfeito	Insatisfeito	Nem Satisfeito Nem Insatisfeito	Satisfeito	Muito Satisfeito
16- Quão satisfeito(a) você está com o seu sono?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
17- Quão satisfeito(a) você está com sua capacidade de desempenhar as atividades do seu dia-a-dia?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
18- Quão satisfeito(a) você está com sua capacidade para o trabalho?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
19- Quão satisfeito(a) você está consigo mesmo?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
20- Quão satisfeito(a) você está com suas relações pessoais (amigos, parentes, conhecidos, colegas)?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
21- Quão satisfeito(a) você está com sua vida sexual?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
22- Quão satisfeito(a) você está com o apoio que você recebe de seus amigos?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
23- Quão satisfeito(a) você está com as condições do local onde mora?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
24- Quão satisfeito(a) você está com o seu acesso aos serviços de saúde?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
25- Quão satisfeito(a) você está com o seu meio de transporte?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

As questões seguintes referem-se a com que freqüência você sentiu ou experimentou certas coisas nas últimas duas semanas.

	Nunca	Algumas Vezes	Freqüente mente	Muito Freqüentemente	Sempre
26- Com que freqüência você tem sentimentos negativos tais como mau humor, desespero, ansiedade, depressão?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Alguém lhe ajudou a preencher este questionário?

Não Sim

Hora de término WHOQOL: :

Tempo total WHOQOL:

AGORA IR PARA E

--	--	--	--

E15: Escolha a alternativa que melhor descreve a sua atividade principal neste momento: **(Leia as opções e marque apenas uma)**

- Desempregado Há quanto tempo?(meses)
- Empregado, com carteira assinada
- Empregado, sem carteira assinada
- Fazendo biscates
- Dona de casa
- Estudante
- Aposentado
- Auxílio doença (encostado por doença)
- Pensionista
- Dono do próprio negócio
- Trabalhador autônomo
- Outro (Especificar):

E16. No último mês, quanto você ganhou? R\$,

(Se não souber ou não quiser responder, complete todas as caselas com 9.)

E17. No último mês, **EXCLUINDO O SR.(A)**, quanto ganharam as pessoas que moram neste domicílio? (considere como renda individual: salários, pensões, bolsa-família, etc - de todos que moram na casa)

Pessoa 1? R\$ por mês ,

Pessoa 2? R\$ por mês ,

Pessoa 3? R\$ por mês ,

Pessoa 4? R\$ por mês ,

Pessoa 5? R\$ por mês ,

Pessoa 6? R\$ por mês ,

E18. No último mês, quanto ganhou o chefe de família? R\$,

(Se não souber ou não quiser responder, complete todas as caselas com 9.)

E19: Sua casa tem água encanada? **(Leia as opções)**

- Sim, dentro da casa Sim, fora da casa Não

E20: Como é a privada da casa? **(Leia as opções)**

- Sanitário com descarga Sanitário sem descarga Casinha Não tem

E21. Quantos banheiros têm na sua casa?

- Nenhum 1 2 3 4 ou mais

--	--	--	--

E22. Quantos dormitórios têm na sua casa?

- Nenhum
 1
 2
 3
 4 ou mais

E23. Vamos fazer perguntas sobre eletrodomésticos e aparelhos que vocês têm em casa. Nós queremos saber se você tem cada um destes aparelhos e quantos destes eletrodomésticos estão funcionando, em condição de uso, esta bem? **(anote somente se estiver funcionando, sempre anotando o nº de artigos - 00 = não tem OU nº de aparelhos caso o entrevistado diga: "sim, tenho")**

Rádio	<input type="text"/>	Máquina de lavar (não considerar tanquinho)	<input type="text"/>
Tv a cores	<input type="text"/>	Forno Microondas	<input type="text"/>
Video/DVD	<input type="text"/>	Linha telefonica	<input type="text"/>
Geladeira	<input type="text"/>	Microcomputador	<input type="text"/>
Frezzer separado da geladeira ou duplex	<input type="text"/>	Condicionador de ar	<input type="text"/>
Aspirador de pó	<input type="text"/>	Automóvel	<input type="text"/>

E24. Você tem empregada mensalista (doméstica, babá, motorista) em sua casa? Se sim, quantas (os)?

- Nenhuma
 1
 2
 3
 4 ou mais

E25. Você tem algum problema de saúde, tais como problema físico, mental ou emocional?

- Sim
 Não
 Não sei/Não lembro

E26. (Se E25=sim) Que problemas de saúde são estes?

Pressão alta (Hipertensão)	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Doença renal	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Diabetes	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Asma/Bronquite	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Cardiopatia isquêmica - angina	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Câncer	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Depressão	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Ansiedade	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Artrose/Artrite	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	HIV/Aids	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Derrame cerebral	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Dor nas costas	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Bronquite crônica/enfisema	<input type="checkbox"/> Sim	<input type="checkbox"/> Não	Insuficiência cardíaca	<input type="checkbox"/> Sim	<input type="checkbox"/> Não
Outras doenças, qual(is):	<input type="text"/>				

--	--	--	--

E27. Quantas vezes o Sr.(a) consultou com médico nos últimos 12 meses?

Nº consultas nos últimos 12 meses

E28. O Sr.(a) esteve hospitalizado nos últimos 12 meses?

Sim Não Não sei/Não lembro

E29. Você fuma cigarros atualmente ou já fumou no passado?

- Sim, fumo atualmente
 Parei de fumar
 Não, nunca fumei

E30. (Se E29=sim) Quantos cigarros você fuma por dia em média? cigarros

E31. Há quanto tempo você parou de fumar? , anos

E32. Nos últimos 30 dias, você consumiu alguma bebida alcoólica?

Sim Não

E33. Nos últimos 7 dias, em quantos dias você ingeriu pelo menos um dose de bebida alcoólica? **(marque 0 caso não tenha consumido um dos tipos)**

- Nº de dias em que consumiu CERVEJA
 Nº de dias em que consumiu VINHO
 Nº de dias em que consumiu DESTILADO

E34. Durante os últimos 30 dias, qual foi a maior quantidade de doses que você ingeriu numa mesma ocasião?

Nº de doses **(1 dose = 1 lata de cerveja ou 1 cálice de vinho ou 1 dose de destilado)**