

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL

CARLA DE BARROS REIS

**Analises custo-utilidade e a comparação de métodos  
alternativos de obtenção de preferências sociais por  
estados de saúde**

Belo Horizonte  
2014

CARLA DE BARROS REIS

# **Analises custo-utilidade e a comparação de métodos alternativos de obtenção de preferências sociais por estados de saúde**

Dissertação de Mestrado apresentada ao Departamento de Pós-Graduação em Economia do Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Economia.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Mônica Viegas Andrade  
Coorientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Kenya Noronha

Belo Horizonte  
2014



## AGRADECIMENTOS

Primeiramente, agradeço a Deus pela saúde, pela paciência e pela energia que me foram concedidas durante todo o período de realização do mestrado. Sem estes três ingredientes fundamentais, eu nada conseguiria alcançar. Agradeço aos meus pais e familiares pelo apoio e carinho nesta etapa tão importante de minha formação. Também dedico sinceros agradecimentos às minhas orientadoras Mônica Viegas Andrade e Kenya Noronha pela confiança que depositaram em mim desde que nos conhecemos na Disciplina Economia da Saúde, em 2010. Receberam-me com todo o carinho no GEESC, em 2011, possibilitando-me travar contato direto com o ambiente de pesquisa acadêmica e adquirir uma rica experiência de aprendizado.

Agradeço aos profissionais responsáveis pela minha formação ao longo do mestrado: professores e funcionários do Cedeplar. Em especial, ao corpo docente da casa, pelos vastos conhecimentos transmitidos e pelo apoio institucional. Agradeço também aos tão queridos professores Ana Flávia Machado, Sueli Moro, Mauro Sayar e Gilberto Libânio, pelo apoio e encorajamento contínuos na pesquisa. Não posso deixar de agradecer também à Maria Cecília e à Cleusa, da secretaria de pós-graduação, que são tão atenciosas com os alunos e nos ajudam sempre que necessário. Obrigada aos meninos da informática pelos momentos de desespero com a nossa “adorável” impressora.

Quero agradecer também a todos os colegas das turmas de mestrado e doutorado, pelas conversas enriquecedoras e pelo compartilhamento de tarefas. Aos colegas da turma de 2012, em especial, à Renata de Miranda Menezes, por ser minha parceira de trabalho. Agradeço a Gustavo Fernandes pelo apoio e pela paciência nos momentos de alegria e desespero. Aos colegas e parceiros da sala 2064, principalmente à Júlia Calazans, pelos bons momentos. Às melhores amigas e pessoas do mundo, Fernanda Alves Cordeiro e Beatriz Judice Magalhães. Vocês tiveram papel decisivo nestes dois anos de mestrado. Obrigada pelos conselhos, paciência e apoio.

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo financiamento recebido ao longo do mestrado e pela possibilidade que oferecem, a milhares de alunos brasileiros, de prosseguirem em seus estudos com alguma dignidade.

## RESUMO

Uma abordagem usual para se medir benefícios em análises de custo-utilidade é o uso dos Anos de Vida Ajustados pela Qualidade - AVAQ, que combina informações sobre mortalidade e qualidade de vida relacionada à saúde. A dimensão de qualidade de vida associada a distintos estados de saúde comumente é medida através de métodos cardinais de obtenção de preferências. Este trabalho tem como objetivos verificar a viabilidade do uso da abordagem ordinal em análises de custo-utilidade e analisar evidências quanto à qualidade e à convergência entre as técnicas de obtenção de preferências populacionais em saúde. Analisa-se em que medida os métodos Ordenação, Escala Analógica Visual e Time Trade-Off são equivalentes, tanto do ponto de vista das utilidades associadas aos estados de saúde, quanto no que se refere ao grau de consistência que apresentam. A base de dados utilizada refere-se a um estudo de avaliação dos parâmetros de valorização dos estados de saúde a partir do EQ-5D, realizado em 2011. Na pesquisa de campo são investigados 3362 indivíduos alfabetizados, com idade entre 18 e 64 anos, residentes de áreas urbanas de Minas Gerais. A amostra probabilística e estratificada por sexo e idade possui uma margem de erro estimada em 3%. A partir do modelo logit condicional e de alguns algoritmos matemáticos, realiza-se a transformação de informações ordinais em dados cardinais e calculam-se as inconsistências lógicas. A comparação entre métodos é verificada a partir de medidas de qualidade de ajuste e coeficientes de correlação. Os resultados apresentados nesta dissertação sugerem a viabilidade do uso de um exercício de avaliação mais simples. O comportamento relativo dos estados de saúde, apresentado pelo método TTO, se mantém no exercício ordinal. Contudo, a utilização deste tipo de abordagem em análises de custo-utilidade ainda apresenta desafios importantes, principalmente, no que se refere à forma de normalização dos pesos. Quanto à qualidade das informações, o Exercício de Troca de Tempo produziu taxas de inconsistência maiores que as demais técnicas, sendo que o nível educacional e a condição socioeconômica foram determinantes importantes na geração de pares inconsistentes. Neste sentido, o exercício de ordenação pode representar uma interessante alternativa às técnicas cardinais de mensuração de preferências por estados de saúde.

**Palavras-chave:** Métodos de avaliação. Ordenação. Comparabilidade. Inconsistências lógicas. EQ-5D. Utilidade.

## ABSTRACT

A common approach to measure benefits in cost-utility analysis is the use of Quality-Adjusted Life Years - QALY , which combines information on mortality and health-related quality of life (HRQoL). The dimension of quality of life associated with different health states has usually been measured by cardinal methods in obtaining preferences. This work aims to verify the feasibility of using the ordinal approach in cost-utility and analyze evidence of the data quality and convergence between the techniques in obtaining societal preferences for health states. We analyze to what extent do the Ranking, Visual Analogue Scale and Time Trade-Off methods yield equivalent results, as of the utilities associated to health conditions as the degree of consistency. Data came from a evaluation study of the parameters for the valuation of health states from the EQ - 5D, carried out in 2011. In the field-work were investigated 3362 literate individuals aged between 18 and 64 years old living in urban areas of Minas Gerais. The probabilistic sample selected on the basis of quota sampling by age and sex has a margin of error equal to 3%. From the conditional logit regression model and some mathematical algorithms perform the transformation of ordinal information to cardinal data and compute the logical inconsistencies. The comparison of methods has been verified from measures of goodness of fit and correlation coefficients. The results suggest the feasibility of using a simple evaluation exercise. The relative tendency of health states presented by the TTO method remains in the ordinal exercise. However the use of this approach in cost-utility analysis also presents significant challenges, especially as the form of values normalization. As regard the data quality, the Time Trade-Off method produced higher inconsistency rate than other techniques and respondents with lower levels of education and socioeconomic status produced larger numbers of inconsistent responses. In this sense, the ranking may to represent an interesting alternative to the cardinal techniques for measuring preferences for health states.

**Keywords:** Valuation methods. Ranking. Comparability. Logical inconsistencies. EQ-5D. Utility.

## LISTA DE FIGURAS

FIGURA 1 – Representação gráfica do <i>Standard Gamble</i> . .....	28
FIGURA 2 – Representação gráfica do <i>Time Trade-Off</i> .....	30
FIGURA 3 – Valores esperados do método da Troca de Tempo e do Exercício de Ordenação segundo diferentes formas de reescalonamento .....	74
FIGURA 4 – Valores observados da EAV e do TTO e valores preditos no modelo logit condicional .....	74
FIGURA 5– Valores preditos do Rank A e observados do TTO por nível de severidade do estado de saúde .....	78
FIGURA 6– Valores preditos do Rank B e observados do TTO por nível de severidade do estado de saúde .....	79
FIGURA 7– Valores preditos do Rank C e observados do TTO por nível de severidade do estado de saúde .....	79
FIGURA 8 – Valores observados da EAV e do TTO por nível de severidade do estado de saúde .....	80
FIGURA 9 – Relação entre os pesos observados e preditos do TTO e do modelo logit condicional segundo três procedimentos de reescalonamento .....	80
FIGURA 10 – Relação entre os pesos observados da EAV e os valores preditos do modelo logit condicional segundo três procedimentos de reescalonamento .....	81
FIGURA 11 – Valores médios observados do TTO segundo nível de inconsistências .....	89
FIGURA 12 – Valores médios observados da EAV segundo nível de inconsistências .....	89
FIGURA 13 – Valores médios preditos do Rank A segundo nível de inconsistências.....	90
FIGURA 14 – Valores médios preditos do Rank B segundo nível de inconsistências.....	90
FIGURA 15 – Valores médios preditos do Rank C segundo nível de inconsistências.....	91
FIGURA A – QUADRO DE TEMPO 1: Estados melhores que a morte .....	113
FIGURA B – QUADRO DE TEMPO 2: Estados piores que a morte .....	113

## LISTA DE TABELAS

TABELA 1 – Sistema Descritivo EuroQol: dimensões de saúde e pontuações do EQ-5D .....	25
TABELA 2 – Características sociodemográficas da amostra e comparação com bases de dados nacionais (%).....	48
TABELA 3 – Descrição e Especificação das Variáveis do Modelo .....	56
TABELA 4 – Descrição e Especificação das Variáveis do modelo.....	70
TABELA 5 – Correlações de <i>Pearson</i> e <i>Spearman</i> dos valores esperados por método .....	73
TABELA 6 – Resultados do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo).....	75
TABELA 7 - Teste de <i>Hausman</i> para a Hipótese de Independência das Alternativas Irrelevantes .....	76
TABELA 8 – Diferença Absoluta Média (DAM) total e por nível de severidade do estado de saúde .....	77
TABELA 9 – Inconsistências lógicas por métodos de avaliação.....	82
TABELA 10 – Percentual de respondentes inconsistentes (%) e distribuição acumulada, por método .....	84
TABELA 11 – Taxa média de inconsistência lógica por características sociodemográficas e de saúde .....	85
TABELA 12 – Razão de chances das inconsistências lógicas por características sociodemográficas e de saúde.....	87
TABELA 13 – Correlação de <i>Pearson</i> entre os pesos da amostra completa e das subamostras definidas a partir do número de inconsistências .....	92
TABELA 14 – Coeficientes estimados a partir do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo) para as amostras completa e consistente .....	93
TABELA A - Valores esperados para os estados de saúde definidos pelo EQ-5D.....	110
TABELA B – Extrato da Matriz de Dominância Lógica .....	111
TABELA C - Resultados do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo).....	112

## LISTA DE ABREVIATURAS E SIGLAS

- AVAQ - Anos de Vida Ajustados pela Qualidade
- CADTH - *Canadian Agency for Drugs and Technologies in Health*
- CCOHTA - Canadian Coordinating Office for Health Technology Assessment
- CEDEPLAR - Centro de Desenvolvimento e Planejamento Regional
- CONITEC - Comissão Nacional de Incorporação de Tecnologias no SUS
- DAM - Diferença Absoluta Média
- DCE - *Discrete Choice Experiments*
- DECIT - Departamento de Ciência e Tecnologia
- EAV - Escala Analógica Visual
- EQ-5D - *EuroQol 5 Dimensions*
- EUROQoL – *European Quality of Life*
- FAPEMIG - Fundação de Amparo à Pesquisa do estado de Minas Gerais
- GEESC - Grupo de Estudos em Economia da Saúde e Criminalidade
- GT/ATS - Grupo Permanente de Trabalho em Avaliações de Tecnologias em Saúde
- HRQoL - *Health-Related Quality of Life*
- HUI - *Health Utility Index*
- IATS - Instituto de Avaliação de Tecnologias em Saúde
- IDH - Índice de Desenvolvimento Humano
- MSAC - *Medical Services Advisory Committee*
- MVH - *Measurement and Valuation of Health Group*
- NHS - *National Health Services*
- NICE - *National Institute for Health and Clinical Excellence*
- OMS - Organização Mundial de Saúde
- PAD - Pesquisa por Amostra de Domicílios
- PBAC - *Pharmaceutical Benefits Advisory Committee*
- PIB – Produto Interno Bruto
- PNAD - Pesquisa Nacional por Amostra de Domicílios
- PNM - Política Nacional de Medicamentos
- PPSUS - Programa Pesquisa Para o SUS
- QALY - *Quality-Adjusted Life Years*
- REBRATS - Rede Brasileira de Avaliação de Tecnologias em Saúde

SF-36 - *Short-Form 36 Items*

SF-6D - *Short-Form 6 Dimension*

SF-12 - *Short-Form 12 Items*

SG - *Standard Gamble*

SUS - Sistema Único de Saúde

TTO - *Time Trade-Off*

UFMG – Universidade Federal de Minas Gerais

UFRGS - Universidade Federal do Rio Grande do Sul

VAS - *Visual Analogue Scale*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	11
<b>2</b>	<b>A TOMADA DE DECISÕES EM POLÍTICAS DE SAÚDE E OS INSTRUMENTOS DE AVALIAÇÃO ECONÔMICA</b> .....	15
<b>2.1</b>	<b>Avaliações de tecnologias em saúde</b> .....	15
<b>2.2</b>	<b>Análises de custo-utilidade e o uso do AVAQ</b> .....	19
<b>2.3</b>	<b>Sistemas Descritivos e o EQ-5D</b> .....	22
<b>3</b>	<b>COMPARAÇÃO DOS MÉTODOS DE OBTENÇÃO DE PREFERÊNCIAS SOCIAIS</b> .....	27
<b>3.1</b>	<b>Medidas cardinais de mensuração de utilidades em saúde</b> .....	27
<b>3.2</b>	<b>A abordagem ordinal como métrica alternativa e a convergência entre os pesos populacionais</b> .....	36
<b>3.3</b>	<b>Inconsistência Interna: o trade-off entre a qualidade das informações e a representatividade amostral</b> .....	39
<b>4</b>	<b>ASPECTOS METODOLÓGICOS GERAIS</b> .....	46
<b>4.1</b>	<b>Base de Dados</b> .....	46
<b>4.2</b>	<b>Transformação de informações ordinais em dados cardinais: o modelo logit condicional</b> .....	51
4.2.1	Exercício de Ordenação .....	52
4.2.2	O reescalamento .....	57
4.2.3	Medidas de Comparação .....	59
<b>4.3</b>	<b>Inconsistências Lógicas</b> .....	61
4.3.1	Definição de Inconsistência .....	62
4.3.2	Cálculo das Inconsistências.....	64
4.3.3	Modelo Logit.....	67
<b>5</b>	<b>RESULTADOS</b> .....	72
<b>5.1</b>	<b>Resultados do Modelo Logit Condicional</b> .....	72
<b>5.2</b>	<b>Resultados das Inconsistências Lógicas</b> .....	81
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	95
	<b>REFERÊNCIAS</b> .....	101
	<b>APÊNDICE A - VALORES ESPERADOS PARA OS ESTADOS EQ-5D</b> .....	110
	<b>APÊNDICE B - EXTRATO DA MATRIZ DE DOMINÂNCIA LÓGICA</b> .....	111
	<b>APÊNDICE C - COMPARAÇÃO ENTRE TTO E RANK C</b> .....	112
	<b>APÊNDICE D - QUADRO DA TROCA DE TEMPO</b> .....	113
	<b>ANEXO A - INSTRUMENTOS</b> .....	114

## 1. INTRODUÇÃO

Tradicionalmente, o envolvimento dos pacientes no processo de tomada de decisões na área médica é mínimo, tanto no nível micro, no atendimento pessoal com o médico; quanto no nível macro, no planejamento e desenvolvimento dos serviços de atenção à saúde. Entretanto, desde a década de 1990, uma maior participação de pacientes e da comunidade nesses processos tem sido estimulada (RYAN e FARRAR, 2000). A obtenção de preferências individuais e comunitárias na área da saúde e sua introdução nas decisões e avaliações de políticas públicas representam um grande avanço neste sentido, já que as utilidades estimadas a partir de pesquisas populacionais permitem alcançar medidas de desfecho da provisão de programas e tratamentos.

As avaliações que ponderam os custos monetários e os benefícios associados à introdução ou substituição de uma tecnologia em saúde, considerando estas preferências sociais, são conhecidas como análises de custo-utilidade. Uma abordagem usual para se medir os benefícios neste tipo de análise é o uso dos Anos de Vida Ajustados pela Qualidade (AVAQ ou QALY do inglês *Quality-Adjusted Life Years*), que combina informações sobre mortalidade e qualidade de vida relacionada à saúde (PHILLIPS, 2009). A dimensão de qualidade de vida associada a distintas condições de saúde pode ser medida através de diversas técnicas, com o objetivo de gerar valores que reflitam a magnitude das preferências individuais.

Metodologias que incorporam a idéia de *trade-off* entre qualidade de vida e longevidade, gerando um resultado cardinal, são comumente utilizadas nas análises econômicas em saúde. Estas técnicas incorporam o conceito microeconômico de utilidade esperada e conduzem a realização de tarefas de difícil administração, com alto nível de abstração e requerimento cognitivo, além da adoção de algumas hipóteses teóricas implícitas na sua construção (LAMERS, 2007; ROBISON, LOOMES e JONES-LEE, 2001; CRAIG, 2009). Alguns métodos cardinais psicométricos apresentam maior facilidade e compreensão por parte dos respondentes, apesar de não se basearem na racionalidade econômica de decisão. Contudo, essas métricas ainda requerem algum grau de abstração, na medida em que os indivíduos devem não somente refletir sobre a descrição dos estados de saúde e a posição relativa que possuem, mas também, atribuir valores numéricos para cada estado de saúde e atentar para a distância relativa entre eles (DOLAN, 2000).

Diferentemente dos métodos cardinais, a abordagem ordinal refere-se aos instrumentos que apresentam apenas uma ordenação dos itens, refletindo as relações de preferência sem qualquer geração de números reais em uma escala intervalar, necessária para a utilização dos AVAQS. No entanto, essas métricas vêm ganhando popularidade, devido sua maior factibilidade e simplicidade na obtenção das preferências com respeito à saúde (SALOMON, 2003; CRAIG, BUSSCHBACH e SALOMON, 2009a; McCABE *et al.*, 2006). Tanto os métodos cardinais quanto os ordinais são baseados nas preferências e se propõem a medir a satisfação dos indivíduos associada a intervenções ou estados de saúde.

Uma maneira de se avaliar diferentes métricas corresponde à verificação do grau em que estes instrumentos mensuram o que se propõem. Normalmente, é estabelecida uma medida de referência com a qual se realizam as comparações (BRAZIER *et al.*, 1999; McDOWELL, 2006). No entanto, o amplo debate sobre qual método deve servir como base para a padronização das avaliações e a ausência de um padrão ouro em medidas de qualidade de vida relacionada à saúde, implicam no estabelecimento de procedimentos indiretos de se analisar esta validade. Um deles refere-se ao exame da equivalência entre os resultados obtidos de diferentes métodos (CRAIG, BUSSCHBACH e SALOMON, 2009b; KRABBE, ESSINKBOT e BONSEL, 1997; BADIA *et al.*, 1999).

Segundo Devlin, Hansen e Selai (2004), a valoração de estados de saúde pode ser considerada como um procedimento complexo de introspecção, avaliação e comparação, fazendo com que os respondentes lidem com situações hipotéticas, muitas vezes nunca experimentadas, e utilizem todas as funções cognitivas, como recepção, memória, aprendizado, pensamento e expressão. Cada um dos métodos destaca, em seus procedimentos, diferentes processos cognitivos, fazendo com que diferenças consideráveis nos resultados sejam observadas (NORD, 1992). Assim, as particularidades de cada métrica, juntamente com o perfil do próprio respondente, impõem consequências importantes no que se refere à representatividade e a qualidade das informações.

De fato, quando se observa o perfil educacional e socioeconômico da população dos países europeus, onde foram pensadas e elaboradas estas metodologias, as dificuldades intrínsecas aos exercícios não estão colocadas em debate ou, pelo menos, não recebem destaque nas análises. Contudo, diante da expansão dos gastos privados e públicos com saúde

experimentada por um número cada vez maior de países e a evidente importância da realização e incorporação de análises de custo-utilidade na alocação de recursos públicos, abre-se espaço para um debate mais crítico com respeito à utilização destes métodos em sociedades mais heterogêneas e com menor nível de desenvolvimento.

A heterogeneidade cultural e socioeconômica se faz fortemente presente na América Latina, principalmente no Brasil. Enquanto no Canadá, Estados Unidos e Alemanha somente 2,4%, 1,7% e 5,0% da população com idade superior a 15 anos possuem apenas o ensino primário completo, no Chile, Argentina e Brasil, estas estimativas chegam a 12,3%, 32,7% e 27,3% em 2010 (THE WORLD BANK, 2013). Em termos de anos médios de estudo, o Brasil apresenta uma média de 7,5 anos de educação formal, seguindo em desvantagem em relação a alguns países latino-americanos (10,2 para Chile e 9,3 para Argentina) e desenvolvidos (12,1, 11,8 e 13,1 para Canadá, Alemanha e Estados Unidos, respectivamente).

Além das discrepâncias internacionais, existem grandes disparidades internas ao país, que se refletem na desigualdade regional, no padrão dual de desenvolvimento socioeconômico representado pelos eixos Centro-Sul e Norte-Nordeste e entre classes econômicas. Neste sentido, questões como grau de abstração e peso cognitivo impostos pelo exercício, dificuldades de administração dos protocolos e custo de pesquisa devem se fazer presentes na escolha das métricas utilizadas. Três métodos de obtenção de preferências foram utilizados na condução de uma pesquisa realizada no Brasil referente às estimações dos parâmetros de valorização dos estados de saúde, conduzida em Minas Gerais a partir do sistema descritivo EQ-5D. A pesquisa de campo foi conduzida em 2011 e os primeiros resultados das estimações dos pesos para o Brasil, utilizando um método tradicional de avaliação, já foram publicados (ANDRADE *et al.*, 2013).

Os estudos a partir das preferências brasileiras abrem espaço para a verificação de evidências quanto à qualidade e convergência entre as medidas de avaliação de estados de saúde para países em desenvolvimento. Além disso, torna-se interessante a discussão de métodos alternativos às abordagens tradicionais, como os exercícios ordinais, que atualmente, possuem pouca relevância prática nos protocolos adotados. Dentro do escopo específico das avaliações econômicas em saúde, a dissertação tem como objetivos verificar a viabilidade do uso da abordagem ordinal em análises de custo-utilidade e analisar em que medida três diferentes métodos de obtenção de preferências populacionais referentes à saúde – Exercício de

Ordenação, Escala Analógica Visual e *Time Trade-Off* - são equivalentes, tanto do ponto de vista das utilidades associadas aos estados de saúde, quanto no que se refere ao grau de consistência que apresentam.

A transformação de informações ordinais em dados cardinais é realizada a partir da regressão logit condicional baseada no modelo de utilidade aleatória. A equivalência entre os diferentes métodos é testada a partir da comparação direta entre os valores observados nos métodos cardinais e os pesos estimados com base nos dados ordinais. No que se refere à consistência interna, avalia-se em que medida as decisões dos respondentes respeitam a ordenação lógica esperada pelo instrumento, refletindo a qualidade das informações segundo a métrica utilizada. Examina-se também a sensibilidade dos pesos obtidos em relação à remoção de respondentes inconsistentes da análise final.

Os resultados apresentados nesta dissertação sugerem a viabilidade do uso de um exercício de avaliação mais simples. O comportamento relativo dos estados de saúde, apresentado pelo método *Time Trade-Off*, se mantém no exercício ordinal. Contudo, a utilização deste tipo de abordagem em análises de custo-utilidade ainda apresenta desafios importantes, principalmente, no que se refere à forma de normalização dos pesos. Quanto à qualidade das informações, o Exercício baseado na troca produziu taxas de inconsistência maiores que as demais técnicas, sendo que o nível educacional e a condição socioeconômica foram determinantes importantes na geração de pares inconsistentes. Neste sentido, o exercício de ordenação pode representar uma interessante alternativa às técnicas cardinais de mensuração de preferências em saúde.

Este trabalho é estruturado da seguinte forma. No capítulo 2, contextualiza-se o objeto de estudo ao apresentar o cenário das avaliações de tecnologias em saúde no Brasil, a descrição dos AVAQs e alguns instrumentos necessários para sua realização. O capítulo 3 apresenta uma revisão bibliográfica sobre a comparabilidade entre os métodos de obtenção de preferências, discutindo-se os métodos e apresentando evidências de equivalência e consistência. A seção 4 descreve a metodologia adotada para a transformação das informações ordinais em dados cardinais, as medidas de comparação entre os pesos e o cálculo das inconsistências lógicas. No capítulo 5, analisam-se os resultados. Por fim, as Considerações Finais apresentam as conclusões e as implicações dos resultados para as políticas públicas.

## **2. A TOMADA DE DECISÕES EM POLÍTICAS DE SAÚDE E OS INSTRUMENTOS DE AVALIAÇÃO ECONÔMICA**

Este capítulo pretende contextualizar o objeto de estudo deste trabalho, apresentando uma figura mais ampla sob a qual as discussões com respeito aos métodos de mensuração de preferências estão inseridas. Esta seção compõe-se de três subseções. Na primeira, apresenta-se o marco legal das avaliações de tecnologias de saúde, além de se discutir os motivos pelos quais estas análises se justificam. Na segunda subseção, descreve-se o avanço na construção de medidas de desfechos em análises econômicas e a importância dos AVAQs como métrica de qualidade de vida. Alguns instrumentos de classificação que descrevem diversos estados de saúde são apresentados na terceira subseção.

### **2.1 Avaliações de tecnologias em saúde**

O dispêndio com cuidados em saúde tem sofrido aumentos significativos nas últimas décadas. Essa tendência crescente dos gastos tem aumentado a preocupação dos gestores de políticas de saúde em alocar de forma mais eficiente os recursos neste setor. De acordo com informações da Organização Mundial de Saúde (OMS), entre 2000 e 2009, os gastos totais com saúde como proporção do PIB aumentaram cerca de 29,5% no Canadá, 29,1% nos Estados Unidos e 40% no Reino Unido. Esse cenário não destoava muito daquele observado em países emergentes, muito embora, nestes países, a participação dos gastos totais em saúde não represente uma parcela expressiva do PIB, especialmente quando comparada com a de países mais desenvolvidos<sup>1</sup>. Entre 2000 e 2009, os gastos totais em saúde como proporção do PIB aumentaram 22,2% no Brasil, 27,5% no México e próximo de 100% em Cuba (WORLD HEALTH ORGANIZATION, 2012).

Alguns fatores relacionados tanto à oferta quanto à demanda por serviços de saúde se apresentam como os principais precursores desta pressão sobre os gastos. Dentre os fatores relacionados à oferta, destaca-se o desenvolvimento de novos equipamentos e medicamentos decorrentes de constantes avanços tecnológicos observados neste setor. A análise da oferta conduz a uma situação ambígua, pois se o aparecimento de novas tecnologias tem fornecido

---

<sup>1</sup> Em 2009, a participação dos gastos totais em saúde no PIB era em torno de 8,8% no Brasil e 6,5% no México contra 17,6% nos Estados Unidos. Dados disponíveis em: [http://www.who.int/gho/publications/world\\_health\\_statistics/EN\\_WHS2012\\_Full.pdf](http://www.who.int/gho/publications/world_health_statistics/EN_WHS2012_Full.pdf).

acesso a novos tratamentos para condições anteriormente consideradas intratáveis, ocasionando melhora na qualidade dos serviços de saúde existentes, também tem propiciado um aumento dos custos com saúde e, conseqüentemente, dos gastos nesse setor (SUEN, 2005; PRODUCTIVITY COMMISSION, 2005). Além disso, a rápida difusão da informação técnico-científica e a atuação de empresas multinacionais criam uma demanda local pela inovação por parte de profissionais da saúde, meios de comunicação e parcelas mais informadas da população, ocasionando uma pressão ainda maior sobre o sistema (BRASIL, 2009a).

No que se refere à demanda, pressões sobre os gastos advêm, principalmente, do envelhecimento populacional. Esse processo tem sido observado em países desenvolvidos e, mais recentemente, em economias em desenvolvimento, sobretudo na América Latina (PALLONI *et al.*, 2006; PALLONI *et al.*, 2005). Segundo informações das Nações Unidas, em países desenvolvidos como um todo, o número de pessoas idosas já ultrapassou o número de crianças e esta distância tende a ser o dobro em 2050. Nos países em desenvolvimento, aproximadamente 8% da população tinha idade superior a 60 anos em 2006 e espera-se que esta proporção chegue a 20% em 2050 (UNITED NATIONS, 2007).

Uma vez que a saúde dos idosos é mais vulnerável que a do restante da população, o processo de envelhecimento populacional tem contribuído para aumentar os gastos. Entre os indivíduos desse grupo etário, a utilização dos serviços de saúde tende a ser mais frequente e os serviços por eles demandados possuem maior conteúdo tecnológico. Além disso, pressões sobre a demanda pelos serviços de saúde podem resultar da melhora na distribuição de renda e do aumento da prevalência e incidência de condições crônicas decorrentes do processo de modernização das sociedades, como obesidade e estresses relacionados a um estilo de vida mais agitado, com conseqüências adversas sobre o estado de saúde (van BAAL *et al.*, 2008).

Nesse cenário de elevação dos gastos com saúde e intenso dinamismo na geração de novas formas de cuidado e inserção de inovações tecnológicas, torna-se fundamental a promoção, de forma sistemática, de avaliações de políticas e tecnologias em saúde que visem uma alocação mais eficiente dos recursos. Através da comparação de custos e benefícios ligados a diferentes formas de intervenções, as avaliações objetivam auxiliar os gestores em suas escolhas, evidenciando possíveis conseqüências da adoção de novas tecnologias ou mudanças em

processos que já estão sendo utilizados (BRASIL, 2008; DRUMMOND, STODDART e TORRANCE, 1997; MURRAY, 2000).

No âmbito internacional, a política de avaliação está consolidada em países como a Austrália, Canadá e Inglaterra, dado que as análises de tecnologias em saúde foram explicitamente incorporadas no processo decisório das políticas públicas destes países (BRASIL, 2008). Um dos países pioneiros no desenvolvimento e utilização deste instrumental foi a Austrália, iniciando esse processo em 1982. O país conta basicamente com duas instituições, *Medical Services Advisory Committee* (MSAC) e *Pharmaceutical Benefits Advisory Committee* (PBAC), responsáveis por avaliar intervenções em saúde e medicamentos. Basicamente, essas instituições avaliam novas tecnologias ou tecnologias já em uso, observando três aspectos: segurança, efetividade e custo-efetividade. Com base nos resultados encontrados, o MSAC e o PBAC são capazes de indicar quais procedimentos em saúde devem ser fornecidos de forma gratuita à população (JACKSON, 2007; NEUMANN, 2005).

No Canadá e na Inglaterra, as avaliações de tecnologias em saúde foram introduzidas no processo decisório através da criação do *Canadian Agency for Drugs and Technologies in Health* (CADTH), em 1989<sup>2</sup>, no Canadá, e do *National Institute for Health and Clinical Excellence* (NICE), em 1999, na Inglaterra. O NICE possui objetivos mais abrangentes, que não se resumem somente ao fornecimento de recomendações ao Sistema Nacional de Saúde ou a prestar informações sobre a efetividade das novas tecnologias neste setor. O instituto também procura estabelecer critérios, de forma a homogeneizar as avaliações, definindo quais segmentos, setores e insumos em saúde são relevantes para essas análises (BRASIL, 2008; NEUMANN, 2005). Os países em desenvolvimento começam a se interessar pelas avaliações econômicas, pois a carência de recursos exige que sejam alcançados resultados mais eficientes (SECOLI *et al.*, 2005).

No Brasil, as avaliações de tecnologias em saúde vêm, paulatinamente, sendo incorporadas nas decisões de alocação de recursos. Essa preocupação vem se tornando cada vez mais notória nos últimos anos, especialmente no que se refere ao financiamento dos medicamentos distribuídos gratuitamente pelo governo. A Política Nacional de Medicamentos – PNM, criada em 1998, possui como principais objetivos: a garantia da segurança, eficácia e

---

<sup>2</sup> Até 2006, o CADTH era conhecido como Canadian Coordinating Office for Health Technology Assessment - CCOHTA (HAILEY, 2007).

qualidade dos medicamentos distribuídos à população, a promoção do uso racional dos medicamentos e o acesso àqueles considerados essenciais. Assim, a PNM impõe regras para a definição dos medicamentos que devem compor a lista daqueles distribuídos gratuitamente pelo governo, priorizando critérios de custo-efetividade (BRASIL, 2000; MOTA, FERNANDES e COELHO, 2003).

Além disso, algumas políticas têm sido consideradas desde 2004 com a criação do Departamento de Ciência e Tecnologia – DECIT, que é responsável por formular e promover avaliações para o Sistema Único de Saúde (SUS). No DECIT desenvolvem-se estudos a partir de importantes diretrizes, tais como o custo social e a gravidade da condição de saúde; a potencial redução dos riscos para saúde e dos custos para o SUS; o custo da tecnologia frente à demanda de utilização; a disponibilidade de evidências científicas; além da pressão realizada por parte das associações de portadores de patologias, pesquisadores, Ministério Público, Judiciário e organismos internacionais (BRASIL, 2006).

Em 2008, a Rede Brasileira de Avaliação de Tecnologias em Saúde – REBRATS foi criada a fim de se estabelecer critérios de priorização mais específicos e diretrizes metodológicas para as avaliações (ANDRADE *et al.*, 2013). Neste sentido, ampliaram-se a integração entre pesquisadores e profissionais da saúde, a disseminação de pesquisas e a qualificação de alguns profissionais das secretarias estaduais e municipais de saúde, através de cursos de especialização e de mestrado profissional. A Comissão Nacional de Incorporação de Tecnologias no SUS – CONITEC (Lei Federal n. 12.401/11), criada em 2011, estabeleceu critérios e prazos para a realização de análises, além de ser responsável pela avaliação de inovações tecnológicas que podem vir a ser utilizadas no sistema público de saúde.

Nos próximos anos, as políticas públicas na área da saúde possivelmente serão influenciadas por estes marcos legais, de modo a permitir avanços nas decisões de alocação de recursos, especialmente, no âmbito de um sistema constitucionalmente universal, como é o caso brasileiro. Devido à escassez de recursos e a crescente intervenção do poder judiciário no setor, a expectativa de obtenção de um sistema capaz de garantir uma assistência integral à população representa um grande desafio aos gestores.

## 2.2 Análises de custo-utilidade e o uso do AVAQ

Os principais métodos utilizados na avaliação econômica em saúde são as análises custo-minimização, análises custo-benefício, análises custo-efetividade e análises custo-utilidade. O que difere uma métrica da outra é a forma com que os benefícios de uma intervenção são definidos (BRASIL, 2008). Os custos, independentemente do tipo de avaliação realizada, são medidos em termos monetários<sup>3</sup>. Nas avaliações de custo-minimização, freqüentemente utilizado para comparar intervenções em saúde que geram o mesmo nível de benefícios, considera-se a intervenção com o menor custo. A principal limitação desse tipo de análise é a impossibilidade de se comparar intervenções que produzam diferentes tipos de desfechos. Na análise custo-benefício os resultados de uma intervenção são definidos em termos monetários. Essa abordagem tem recebido críticas relacionadas, principalmente, à questões éticas, tanto por envolver uma valorização da vida humana em termos monetários, quanto por conferir menor peso a condições específicas de trabalhadores cuja taxa salarial é mais baixa (HURLEY, 2000).

A análise custo-efetividade mede os benefícios em unidades naturais dos resultados esperados de uma intervenção, como por exemplo, número de anos de vida salvos, número de casos evitados, número de internações hospitalares evitadas. Portanto, a unidade em que os benefícios são mensurados depende da natureza da intervenção avaliada. A principal vantagem dessa metodologia é que os resultados são de fácil compreensão, podendo ser diretamente interpretados por gestores, médicos e outros profissionais da área da saúde. Em contrapartida, além de não permitir comparação de intervenções de saúde que gerem diferentes resultados, não é possível inserir na análise mais de um desfecho associado a uma mesma intervenção (HURLEY, 2000; GOLD, FRANKS e ERICKSON, 1996; DRUMMOND *et al.*, 1997).

Com relação ao indicador que mede o número de anos de vida salvos ou número de anos de vida ganhos decorrente de uma intervenção, maior peso é atribuído a tecnologias de saúde direcionadas a doenças cuja letalidade é mais alta por não incorporar impactos sobre reduções

---

<sup>3</sup> Os custos são compostos por custos tangíveis e intangíveis. Os custos tangíveis são mensurados através do valor de mercado associado. Eles podem ser diretos (relacionados com a intervenção em saúde) ou indiretos (custos de produtividade estimados em termos do valor médio do tempo de trabalho perdido). Os custos intangíveis não possuem um preço de mercado associado, como por exemplo, dor, tristeza, angústia, ansiedade (SECOLI *et al.*, 2005).

em morbidade. Em países que possuem um perfil epidemiológico cuja letalidade das doenças ainda é muito elevada, essa limitação pode apresentar menor relevância uma vez que os benefícios de diferentes intervenções serão compostos principalmente por ganhos em longevidade. Nas economias desenvolvidas e em desenvolvimento, a limitação desse indicador passa a ter maior importância. Nesses países, as avaliações econômicas em saúde frequentemente objetivam comparar tecnologias que são direcionadas à prevenção e ao tratamento de doenças com níveis distintos de letalidade (ROBBERSTAD, 2005).

No Brasil essa questão é especialmente relevante, na medida em que coexistem no país diferentes perfis epidemiológicos. Se por um lado, nas regiões mais desenvolvidas, como o Sul e Sudeste, o perfil epidemiológico é caracterizado pelo predomínio de doenças crônico-degenerativas que trazem algum grau de prejuízo à qualidade de vida, por outro lado, nas regiões mais pobres, como o Norte e Nordeste do país, ainda permanecem, com relativa importância, doenças relacionadas à pobreza, como as infecto-contagiosas e parasitárias, subnutrição e taxas relativamente elevadas de mortalidade, sobretudo infantil (SCHRAMM *et al.*, 2004). Com o processo de envelhecimento populacional e o correspondente aumento de condições crônicas no país, o uso de indicadores tradicionais de saúde, tais como taxa de mortalidade e expectativa de vida, não é suficiente para o monitoramento da saúde da população brasileira.

Uma alternativa para superar esta limitação é a realização de avaliações econômicas baseadas na **análise custo-utilidade**. Essa técnica combina informações sobre mortalidade com uma medida de qualidade de vida relacionada à saúde (HRQoL do inglês *Health-Related Quality of Life*). Qualidade de vida é definida como “a percepção dos indivíduos sobre a sua posição na vida, no contexto da sua cultura e sistema de valores, em relação aos seus objetivos, expectativas e preocupações” (BRASIL, 2009b). No campo da saúde, entretanto, é preferível restringir a abordagem à qualidade de vida relacionada à saúde, considerando apenas os aspectos relativos à saúde do paciente. A HRQoL permite que critérios subjetivos (físicos, sociais e emocionais) sejam combinados aos critérios de avaliação puramente clínica.

Neste sentido, Klarman *et al.* (1968) desenvolveu o conceito de Anos de Vida Ajustados pela Qualidade (AVAQ ou QALY do inglês *Quality-Adjusted Life Years*). O AVAQ é uma medida única composta pelo número de anos de vida, ponderado por um índice resumo, ou peso, relacionado à qualidade de vida associada à determinada condição de saúde. Assim, o uso do

AVAQ é apropriado quando a qualidade de vida é um importante resultado ou quando o programa sob avaliação afeta ambos, morbidade e mortalidade, desejando-se obter uma unidade comum de benefício que combina ambos os efeitos. Alguns estudos sobre longevidade e saúde mostram que ganhos em expectativa de vida têm sido acompanhados por um aumento dos anos de vida experimentando alguma doença crônica ou outras morbidades (ROMERO, LEITE e SZWARCOWALD, 2005; CAMARGOS, PERPÉTUO e MACHADO, 2005).

A idéia subentendida nesse conceito é que existe um *trade-off* entre longevidade e qualidade de vida. Neste sentido, as análises de custo-utilidade consideram o custo monetário de uma determinada intervenção, e no cálculo da unidade de benefício, encontra-se implícita a idéia de utilidade esperada. Estes resultados reforçam o debate sobre o uso indiscriminado de novas tecnologias que somente afetam a extensão do tempo de vida e revelam o avanço conquistado com a incorporação dos AVAQs nas avaliações econômicas em saúde<sup>4</sup>. Na presença de determinadas doenças, por exemplo, o tratamento possui um efeito inicial positivo, em termos de anos de vida ajustados à qualidade, mas com o decorrer do processo, impõe efeitos colaterais, que diminuem o benefício inicial. Nesta etapa, a qualidade de vida do paciente pode tornar-se inferior à qualidade de vida associada ao indivíduo que não se expôs ao tratamento. A partir de determinado ponto no tempo, os indivíduos que não se tratam entram em óbito, enquanto os demais experimentam ganhos em AVAQs, em virtude da extensão da longevidade, mesmo que em um nível de qualidade de vida inferior àquela associada a indivíduos saudáveis.

A comparação entre duas intervenções alternativas também pode ser realizada a partir dos AVAQs. Por exemplo, um ano em saúde perfeita, na qual a qualidade de vida é plena (peso igual a 1), refere-se a um AVAQ. Já um ano sob um estado de saúde com determinada morbidade, avaliado em 0.5, ou seja, uma condição de saúde que apresenta um decréscimo na qualidade de vida em 50% relativo à saúde perfeita, é equivalente à metade de um AVAQ. Assim, uma intervenção que gera 4 anos de vida adicionais em um determinado estado de saúde avaliado em 0.75 produz um AVAQ a mais que uma intervenção que gera 4 anos de vida adicionais em um

---

<sup>4</sup> Apesar dos avanços conquistados com a incorporação dos AVAQs, esta medida traz algumas hipóteses restritivas (LOOMES e McKENZIE, 1989). O princípio ético inerente a este método é que um ano de vida saudável vale 1, independentemente do indivíduo. Isto significa que o valor do AVAQ independe das características socioeconômicas e demográficas do indivíduo. Por exemplo, um AVAQ de um indivíduo de 25 anos é equivalente a um AVAQ de um indivíduo de 85 anos de idade. Outro princípio fundamental deste método é que **um AVAQ ganho é um AVAQ ganho**, independentemente da forma como foi obtido. Desse modo, o ganho em qualidade de 0.5 de um indivíduo é equivalente ao ganho em qualidade de 0.25 de dois indivíduos, que é equivalente ao ganho em qualidade de 0.1 de cinco indivíduos, e assim por diante. Analogamente, um ganho em qualidade de 0.5 por um ano é equivalente a um ganho em qualidade de 0.25 por dois anos. Assim, prevalece a questão da eficiência versus equidade, na medida em que as decisões médicas não podem considerar apenas as questões relativas à alocação de recursos, mas também observar a justiça distributiva.

determinado estado de saúde avaliado em 0.50 (PHILLIPS, 2009). Portanto, o AVAQ é o produto entre a expectativa de sobrevida e uma medida de qualidade dos anos de vida ganhos. Enquanto os conceitos de sobrevida e expectativa de vida são tradicionais e não implicam em problemas de comparação, a qualidade de vida relacionada à saúde engloba diferentes dimensões, como capacidade mental, física e emocional. Diversas abordagens têm sido usadas para gerar as avaliações referentes à qualidade de vida. A seguir são apresentados alguns sistemas descritivos, que oferecem ampla diversidade de condições ou estados de saúde.

### **2.3 Sistemas descritivos e o EQ-5D**

Muitos tratamentos, produtos e tecnologias na área da saúde proporcionam, à população atingida por certas doenças, vários estados de saúde possíveis. Baseadas em amostras representativas da população, as pesquisas de avaliação custo-utilidade objetivam captar como os indivíduos valorizam estes estados de saúde a fim de se estimar, para cada um deles, um parâmetro médio. Assim, o AVAQ requer a construção de pesos para cada estado de saúde, que são calculados a partir da valorização média da população associada ao tratamento de cada morbidade. A percepção individual do estado de saúde geralmente é avaliada em diferentes domínios da vida, ou seja, em determinadas áreas do comportamento ou da condição humana, como as dimensões física, psicológica, econômica, espiritual e social.

Os instrumentos de avaliação de qualidade de vida que abarcam esses domínios podem ser genéricos ou específicos. Os primeiros são aplicáveis em diferentes contextos e intervenções em saúde, já que possibilitam uma avaliação simultânea de várias áreas ou dimensões, podem ser utilizados em qualquer população e permitem comparações entre pacientes com diferentes patologias e morbidades. Já os sistemas específicos são designados a condições específicas de doença (diabetes, artrite, doença pulmonar, por exemplo), funções (função sexual, por exemplo) ou populações (crianças, idosos, adultos), e possuem maior sensibilidade clínica aos domínios de relevância do aspecto a ser avaliado (CAMPOLINA e CICONELLI, 2006).

Existem diversos sistemas de classificação que podem ser usados na construção dos AVAQs, dentre os quais destacam-se o *Health Utility Index* (HUI), o *Short-Form 36 Items* (SF-36) e sua simplificação - o *Short-Form 6 Dimension* (SF-6D), além do *EuroQol 5 Dimensions* (EQ-5D). Basicamente, estes instrumentos se diferenciam pelo número e tipo de dimensões consideradas, além dos níveis de severidade pertencentes em cada domínio. O *Health Utility*

*Index* compreende três sistemas descritivos – HUI 1, HUI 2 e HUI 3, que foram desenvolvidos ao longo do tempo, e cada um baseia-se no anterior. Designado para *surveys* populacionais, o sistema mais completo em termos de atributos, o *Health Utility Index III* (HUI III), foi desenvolvido no Canadá e é composto de oito atributos, quais sejam: visão, audição, fala, mobilidade, destreza, emoção, cognição e dor/desconforto. Cada dimensão abrange cinco ou seis níveis de severidade, definindo um total de 972 mil estados de saúde teóricos (GOLD, FRANKS e ERICKSON, 1996; FEENY *et al.*, 2002).

Mais recentemente, várias tentativas de cálculo de utilidades têm sido feitas a partir de dados levantados com o SF-36, um instrumento genérico de qualidade de vida composto de 36 itens, que mede a saúde através de oito atributos: funcionamento físico, peso das limitações devido aos problemas físicos, funcionamento social, dor corporal, peso das limitações devido a problemas emocionais, saúde mental, vitalidade e percepção geral da saúde (BRAZIER *et al.*, 1992; BRAZIER, ROBERTS e DEVERILL, 2002; HURST, RUTA e KIND, 1998). O SF-12 e o SF-6D foram desenvolvidos como alternativas sintéticas ao SF-36, contendo menor número de itens para cada uma das oito dimensões identificadas no instrumento original (BRAZIER *et al.*, 2004; CUNILLERA *et al.*, 2010).

O instrumento EuroQol 5 Dimensões (EQ-5D), desenvolvido por um grupo de especialistas da Universidade de York (*Measurement and Valuation of Health Group* – MVH) (GUDEX, 1994), é a medida genérica mais utilizada para se medir benefícios de saúde em avaliações econômicas. O instrumento é recomendado pelo *National Institute for Health and Clinical Excellence* (NICE), que é responsável por desenvolver guias para análises de tecnologias no âmbito do Serviço Nacional de Saúde do Reino Unido - *National Health Services* (NHS) (NICE, 2008). De fato, trata-se de uma abordagem que captura em seus domínios, as condições físicas, mentais e sociais do grupo alvo. Mais do que sua utilização em avaliações para uma determinada doença ou tratamento específico, o instrumento é designado para o uso em um intervalo amplo de intervenções em saúde.

O Sistema Descritivo EQ-5D consiste em uma métrica que define qualidade de vida relacionada à saúde em termos de cinco dimensões (mobilidade, autocuidado, atividades usuais, dor/desconforto, ansiedade/depressão) com três níveis cada uma (sem problemas, alguns problemas e problemas extremos), apresentando a descrição de 243 distintos estados de saúde ( $3^5$ ). Além disso, o estado ‘morte’ também é incorporado nesta estrutura. A fim de

estimar a métrica baseada nesse questionário, em 1993, o grupo MVH realizou uma pesquisa na Inglaterra investigando 42 estados de saúde definidos a partir do EQ-5D (DOLAN *et al.*, 1996a). A partir da avaliação direta realizada para estes estados de saúde, utilizaram-se métodos de regressão para interpolar os valores dos 200 estados restantes. Os resultados encontrados pelo MVH são amplamente utilizados para avaliações de custo-utilidade em diversos países, sobretudo os europeus.

A TABELA 1 abaixo apresenta a pontuação para cada nível e dimensão do sistema descritivo. Quando o respondente seleciona um nível em cada dimensão, ele está definindo seu próprio estado de saúde em termos de um código de cinco dígitos. Assim, o estado de saúde 12113, por exemplo, reflete ausência de problemas na mobilidade, atividades usuais e dor/desconforto, alguns problemas na dimensão de autocuidado e problemas extremos de ansiedade/depressão.

No que se refere à convergência dos sistemas descritivos, os resultados encontrados apresentam algumas evidências interessantes. Macran, Weatherly e Kind (2003) sugerem que, a despeito das diferenças inerentes entre as medidas descritivas, principalmente em termos de sua apresentação, número de itens e diversidade de categorias, as performances do EQ-5D, HUI3 e SF-12 são muito similares. Em termos de viabilidade prática, os três sistemas apresentaram baixo (menos que 6%) percentual de indivíduos que não conseguiram realizar as tarefas. O EQ-5D e o SF-12 são amplamente comparáveis no que se refere à discriminação de subgrupos populacionais quanto à suas características sociodemográficas, econômicas, condições crônicas e problemas recentes de saúde. O SF-6D parece ser menos sensível às condições crônicas físicas que o EQ-5D e o SF12 (CUNILLERA *et al.*, 2010).

Brazier *et al.* (2004) encontram ampla convergência entre dimensões similares do EQ-5D e do SF-6D. Contudo, o primeiro apresenta grande proporção de respondentes que se auto-avaliam na melhor categoria de cada dimensão, sugerindo uma menor capacidade em distinguir estados de saúde próximos à saúde perfeita, quando comparado ao segundo. Assim, pacientes que relatam possuir saúde perfeita no sistema EQ-5D, podem estar sofrendo com problemas no funcionamento físico, saúde mental e vitalidade. Além disso, o SF-6D gera maiores valores médios e menores valores medianos para todos os estados de saúde. Portanto, a própria definição do sistema de classificação usado para avaliar qualidade de vida pode gerar

impactos nas avaliações de saúde, sendo necessárias análises de sensibilidade destas variações sobre as escolhas de intervenções e programas.

**TABELA 1 – Sistema Descritivo EuroQol: dimensões de saúde e pontuações do EQ-5D**

<b>Dimensão</b>	<b>Nível</b>	<b>Pontuação</b>
Mobilidade	Não tenho problemas para andar	1
	Tenho alguns problemas para 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 vestir sozinho	3
Atividades Habituais	Não tenho problemas para desempenhar minhas atividades habituais	1
	Tenho alguns problemas para desempenhar minhas atividades habituais	2
	Sou incapaz de desempenhar minhas atividades habituais	3
Dor/Desconforto	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

Fonte: Adaptação para o português de Dolan e Roberts (2002).

Para que o EQ-5D e demais sistemas descritivos sejam utilizados em avaliações de benefícios em saúde associados com diferentes intervenções neste tipo de cuidado, é necessário derivar uma medida resumo, um índice para cada um desses estados. A obtenção destes pesos sociais associados a diversas condições de saúde, necessários para a construção dos AVAQs, é realizada através de pesquisas aplicadas em amostras capazes de representar a população em geral ou aplicadas a pacientes com perfis específicos de saúde. Estimativas destas tarifas estão disponíveis para diversos países, especialmente em economias mais desenvolvidas tais como Austrália, Canadá, Inglaterra, Holanda, Espanha, Alemanha e Japão (DOLAN *et al.*, 1996a; CHEVALIER e POUVOURVILLE, 2013; BANSBACK *et al.*, 2012a; TSUCHIYA *et al.*, 2002). Recentemente, somente três países da América Latina - Argentina, Chile e Brasil - realizaram surveys populacionais para a obtenção do conjunto de valores que representam as preferências nacionais por estados de saúde (AUGUSTOVSKI *et al.*, 2009; ZARATE *et al.*, 2011; ANDRADE *et al.*, 2013).

A geração destes pesos pode ser realizada através de diferentes métodos. Estas técnicas determinam utilidades individuais em uma escala comum, que são agregadas a partir de uma medida de tendência central, formando a utilidade social relacionada a uma determinada

condição de saúde. A mensuração de preferências ordinais verifica-se trivial, dado que um posicionamento dos estados de saúde segundo a ordem de preferência do indivíduo possibilita alcançar a satisfação relativa entre os itens avaliados. Contudo, a abordagem ordinal impõe dificuldades na agregação de utilidades individuais e na geração de pesos necessários para as análises de custo-utilidade, na medida em que não estabelece estimativas diretas sobre o nível em que um estado de saúde é preferível a outro (ALI e RONALDSON, 2012).

Neste sentido, as avaliações econômicas utilizam medidas cardinais de preferências, pois estas oferecem um conjunto de valores na reta real para os resultados de cada estado de saúde, adotando-se uma escala intervalar. Além disso, as métricas cardinais diferenciam os estados de saúde de forma mais precisa, na medida em que permitem uma gama de valores para cada uma das condições (CRAIG, BUSSCHABACH e SALOMON, 2009a). As principais técnicas cardinais utilizadas para a obtenção dos pesos são a Escala Analógica Visual – EAV (do inglês *Visual Analogue Scale* - VAS), o Exercício da Troca de Tempo (do inglês *Time Trade-Off* - TTO) e a Loteria Padrão (do inglês *Standard Gamble* - SG). As descrições das métricas e suas potenciais vantagens e desvantagens são discutidas na próxima seção.

### 3. COMPARAÇÃO DOS MÉTODOS DE OBTENÇÃO DE PREFERÊNCIAS SOCIAIS

Este capítulo apresenta uma revisão da literatura com respeito às comparações entre diferentes métricas de avaliação e compõe-se de três subseções. Na primeira, descrevem-se os métodos cardinais de obtenção de preferências por estados de saúde, que correspondem às métricas tradicionais comumente utilizadas em análises de custo-utilidade. Na segunda subseção, discute-se a abordagem ordinal e sua potencial utilização em avaliações econômicas do setor. Na terceira subseção analisa-se o fenômeno das inconsistências lógicas, que apresenta importantes implicações para a qualidade dos dados utilizados na elaboração de políticas públicas na área da saúde.

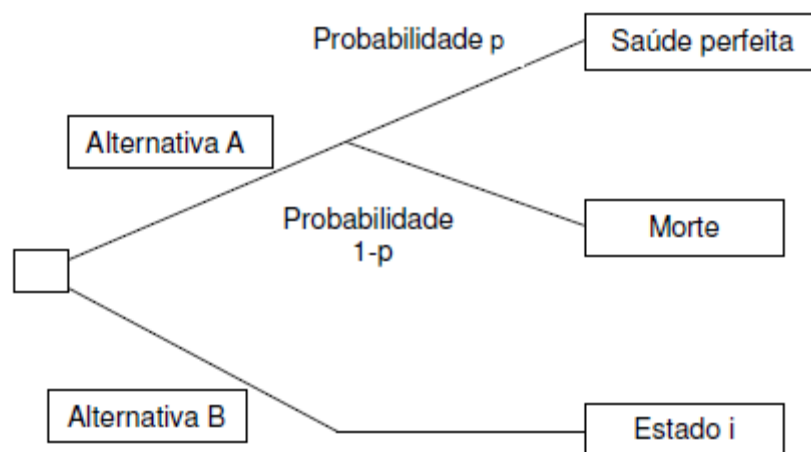
#### 3.1 Medidas cardinais de mensuração de utilidades em saúde

O *Standard Gamble* é o método mais tradicional em avaliações econômicas em saúde, pois sua base conceitual é derivada da teoria da utilidade esperada, descrita por 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, STODDART e TORRANCE, 1997). A abordagem consiste em derivar uma função de utilidade para o indivíduo baseada em suas preferências diante de uma loteria. A fim de que a função utilidade seja capaz de representar essas preferências algumas suposições são necessárias. Os indivíduos devem possuir uma relação de preferência racional, ou seja, deve ser completa e transitiva, além de respeitar o axioma da independência.

As relações de preferências são ditas completas se é possível estabelecer uma ordem de preferência entre cada par de desfechos. Assim, em um par de desfechos, um indivíduo prefere um dos dois ou é indiferente a eles. A transitividade requer consistência na relação de preferências, de modo que se um desfecho A é preferido ao desfecho B, que por sua vez, é preferido ao desfecho C, então A é preferido a C. O axioma da independência assume que a relação de preferências entre duas loterias se mantém mesmo com a introdução de uma terceira. Assim, se A é preferido a B, então  $A + C$  é preferido a  $B + C$ . No contexto das avaliações em saúde, independência significa que a utilidade atribuída a uma condição de saúde particular não deve ser influenciada por resultados alternativos oferecidos no jogo, ou seja, o valor de um dado resultado é independente do seu contexto (BRAZIER *et al.*, 1999).

No método *Standard Gamble*, o indivíduo realiza uma escolha entre resultados alternativos, no qual em um deles existe um componente de incerteza. Ao sujeito são oferecidas duas alternativas. A opção A refere-se a um tratamento com dois resultados possíveis: com probabilidade  $p$ , o paciente retorna à saúde perfeita e permanece neste estado por  $t$  anos adicionais; ou o paciente morre de forma imediata com probabilidade  $1-p$ . Assim, representando a função de utilidade individual como  $V = p.h_i$ , onde  $h_i$  refere-se a condição de saúde  $i$ , a utilidade esperada nesta aposta corresponde a  $V_A = p.1 + (1-p).0 = p$ . A opção B apresenta um resultado certo: o indivíduo vive por  $t$  anos sob a condição crônica de saúde. A utilidade esperada nesta aposta corresponde a  $V_B = 1.h_i = h_i$ .

A probabilidade  $p$  sofre alterações até que o respondente seja indiferente entre as duas alternativas. Assim, a curva de indiferença do indivíduo é composta pela igualdade entre as utilidades derivadas nas duas alternativas, ou seja,  $V_A = V_B \leftrightarrow h_i = p$ . Logo, a utilidade do estado de saúde avaliado é igual a probabilidade  $p$ . A probabilidade de sucesso associada à indiferença do respondente entre duas alternativas provê uma unidade de medida para o valor do estado de saúde. A FIGURA 1 apresenta a representação gráfica da técnica. Se o estado de saúde for muito indesejável, o indivíduo aceita uma intervenção para alterar sua condição mesmo com uma baixa probabilidade de recuperar sua saúde normal. Neste caso, a utilidade associada a este estado de saúde é baixa.



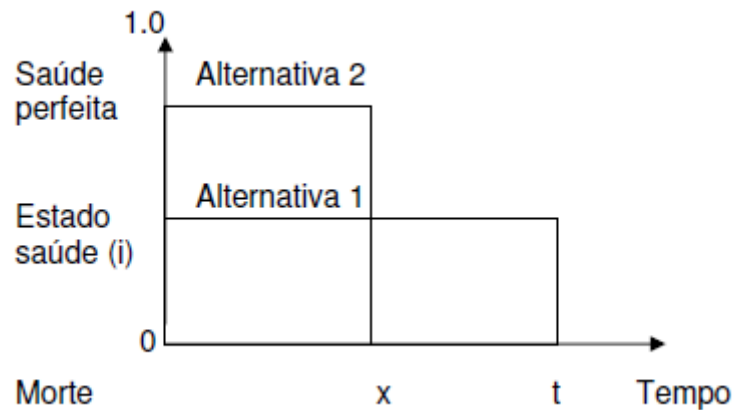
**FIGURA 1 – Representação gráfica do *Standard Gamble*.**

Fonte: Cruz (2010).

Para estados considerados piores que a morte, a alternativa sob incerteza oferece ao indivíduo um resultado de saúde perfeita com probabilidade  $p$  ou uma determinada condição de saúde com probabilidade  $1-p$ , enquanto a alternativa sob certeza refere-se à morte. Assim, as utilidades associadas às alternativas são  $V_A = p \cdot 1 + (1-p) \cdot h_i$  e  $V_B = 1 \cdot 0 = 0$ . Neste caso a utilidade do estado de saúde avaliado  $h_i$  é igual a  $-p/(1-p)$ . Neste caso, se um estado de saúde é extremamente indesejável, um indivíduo provavelmente não escolheria a aposta, a não ser que a probabilidade de retornar ao estado de saúde normal fosse alta. Estados piores que a morte são representados por números negativos. O método sempre envolve uma escolha entre um desfecho de certeza e outro de incerteza, onde o estado de saúde associado ao desfecho de ocorrência certa é intermediário em termos de preferência em relação ao melhor e pior estado. Como a técnica baseia-se na tomada de decisão sob um ambiente de incerteza, respeitando o referencial teórico da Teoria da Utilidade Esperada, o uso do *Standard Gamble* é considerado ideal em análises econômicas em saúde.

O método da Troca de Tempo, desenvolvido por Torrance, Thomas e Sackett (1972) como uma alternativa ao SG, supera as dificuldades relacionadas à necessidade de se explicar o conceito de probabilidade aos respondentes. De acordo com essa metodologia, os indivíduos são chamados a escolher entre duas alternativas que envolvem uma troca entre quantidade e qualidade de vida, sob cenários de certeza. Ao sujeito são oferecidas duas alternativas. Na primeira opção, o indivíduo vive sob determinada condição de saúde por  $t$  anos seguidos de morte, enquanto na segunda opção, o respondente vive por  $x < t$  anos sob saúde perfeita, seguidos por morte. O período  $x$  sofre alterações até que o respondente seja indiferente entre as duas alternativas.

A FIGURA 2 ilustra a técnica. Quanto melhor for o estado de saúde avaliado, menos tempo de vida o indivíduo está disposto a abrir mão a fim de se obter ganhos em qualidade de vida. Assim, os respondentes permitem reduções em termos de longevidade por melhoras na qualidade de vida. Representando a função de utilidade individual como  $V = T \cdot h_i$ , onde  $T$  corresponde aos anos de vida, o valor da utilidade corresponde à fração de ano de vida com saúde plena equivalente a um ano de vida sob uma determinada condição ou estado de saúde, ou seja,  $h_i = x/t$  (MEHREZ e GAFNI, 1990).



**FIGURA 2 – Representação gráfica do *Time Trade-Off***

Fonte: Cruz (2010).

De forma semelhante ao SG, o TTO foi adaptado para avaliar estados de saúde considerados como piores que a morte. Neste caso, a alternativa sob certeza envolve a morte imediata, enquanto a alternativa sob incerteza envolve a escolha entre x anos na condição de saúde classificada como pior que a morte, seguida por t-x anos em saúde perfeita, como recompensa. Mais uma vez, a duração x é modificada até que o indivíduo seja indiferente entre as duas alternativas. O valor do estado de saúde avaliado corresponde à razão  $-x/(t-x)$ . Cada indivíduo valoriza alguns estados de saúde através de um quadro de dupla-face, no qual em um dos lados se avaliam estados considerados pelos respondentes como melhores que a morte e, na outra face, estados registrados como piores que a morte.

Portanto, o exercício da Troca de Tempo identifica diferentes pontos da curva de indiferença individual, possibilitando uma troca entre ganhos em quantidade e qualidade de vida, sob o arcabouço da teoria da utilidade esperada sob um ambiente de certeza. Por possuir a idéia de sacrifício ou troca entre argumentos da função utilidade individual e representar uma métrica mais factível que o SG, o TTO é a técnica mais utilizada para a obtenção dos pesos para a construção dos AVAQS. Como os métodos *Time Trade-Off* e *Standard Gamble* derivam os valores das preferências baseando-se nas respostas dos indivíduos em situações de decisão, estas métricas são denominadas *choice-based* e são preferidas em avaliações econômicas em saúde por permitir estimar a variação de bem-estar associada com uma mudança na condição de saúde e a variação compensatória em outros argumentos necessária para manter a utilidade constante (DOLAN, 2000).

Contudo, os métodos *choice-based* apresentam-se como métricas que exigem maior grau de abstração e peso cognitivo dos participantes (COAST *et al.*, 2008). Além disso, se os axiomas necessários para a manutenção da Teoria da Utilidade Esperada forem violados, a justificativa para preferir os métodos baseados na troca, com base na sua superioridade teórica, deve ser questionada (BRAZIER *et al.*, 1999). Diversas evidências de violação dos axiomas da teoria microeconômica de decisão, no contexto da saúde, são reportadas por Dolan (1996), Read *et al.*, (1984), Stiggelbout *et al.*, (1996), Dolan (2000) e Loomes e McKenzie (1989).

No caso do TTO, por exemplo, o axioma da independência está intimamente ligado ao efeito da duração no estado de saúde sobre a avaliação. Uma hipótese do método é que os indivíduos estão dispostos a trocar uma proporção constante de tempo de vida para melhorar a qualidade de vida, independentemente do número de anos restantes. Contudo, a avaliação de um estado de saúde pode ser influenciada pelo tempo que um indivíduo pode permanecer hipoteticamente neste estado. Indivíduos podem adaptar-se ou tornarem-se intolerantes a determinadas condições de saúde, ao longo do tempo. Dessa forma, o uso de um período de duração específico – 10 anos, por exemplo – pode afetar as utilidades obtidas a partir do TTO (BRAZIER *et al.*, 1999).

Bleichrodt e Johannesson (1996) testam a hipótese de que a utilidade derivada do tempo de vida é independente da utilidade associada à qualidade de vida. Em uma amostra de 172 estudantes, os autores verificam que menos de 15% dos respondentes satisfazem a condição de independência das utilidades. Cerca de 75% dos indivíduos atribuem valores mais baixos para estados de saúde com duração de 10 anos, quando comparado a um período de 30 anos. Segundo Sackett e Torrance (1978), os indivíduos estão dispostos a trocar uma maior fração de quantidade por qualidade à medida que se eleva o tempo de permanência na condição de saúde. Craig (2009) testa a hipótese de proporcionalidade constante a partir do estudo nacional para a população do Reino Unido. Os autores sugerem uma relação positiva entre os valores dos estados de saúde precários e o tempo de duração. Estes resultados corroboram o fenômeno da adaptação dos indivíduos a piores condições de saúde.

A preferência intertemporal nas avaliações do TTO está intimamente ligada à hipótese da proporcionalidade constante das trocas e refere-se à importância relativa que a ocorrência de um estado de saúde possui a depender do período em que é experimentado. Isto pode

contradizer a hipótese de proporcionalidade constante, já que se um indivíduo possui uma taxa de preferência de tempo positiva, atribuirá maior valor aos anos de vida num futuro próximo que os anos de vida em um futuro distante. Dolan e Gudex (1995) suportam a evidencia do efeito da preferência temporal e da violação da hipótese de proporcionalidade constante das trocas. A maioria das respostas implicou em uma taxa negativa.

Ainda no que tange ao método da Troca de Tempo, uma questão que tem merecido pouco destaque na literatura de avaliação refere-se à assimetria das distribuições de valores para estados de saúde classificados como piores e melhores que a morte. Para estados classificados como melhores que a morte, as utilidades são positivas com valor máximo igual a 1, que equivale à saúde perfeita. Por definição, o estado Morte possui utilidade nula. Para estados avaliados como piores que a morte, a utilidade é negativa. Contudo, neste último caso, não existe um limite inferior teórico para a utilidade destes estados, atribuindo um maior peso para os valores negativos no cálculo das utilidades médias (LAMERS, 2007; DOLAN *et al.*, 1996a). Recomenda-se que os estados classificados como piores que a morte sejam limitados inferiormente pelo valor -1, trabalhando-se com um intervalo equidistante da morte em ambas direções, positiva e negativa. A literatura apresenta diversos métodos para se limitar os valores de estados piores que a morte. Contudo, Lamers (2007) mostra que a escolha do método é arbitrária e afeta o conjunto de pesos finais.

Um outro método de obtenção de preferências é a Escala Analógica Visual, que foi originalmente desenvolvida no contexto da psicometria, a fim de se mensurar a sensibilidade subjetiva dos indivíduos com respeito a estímulos sensoriais, tais como luz, calor e som. Esta métrica consiste em uma escala, tipicamente representada por um termômetro vertical, com extremidades 0 e 100 (KRABBE *et al.*, 2006). Em muitos casos, estas extremidades correspondem à saúde perfeita e à morte, respectivamente. Para permitir a possibilidade que alguns estados sejam classificados como piores que a morte, alguns sistemas de classificação renomeiam estas extremidades como melhor estado de saúde imaginável e pior estado de saúde imaginável. Os estados de saúde são posicionados entre as duas extremidades, segundo as preferências do respondente, tal que o intervalo entre dois estados quaisquer corresponda à diferença entre as preferências percebidas pelo indivíduo por ambos estados de saúde.

Apesar de ser aceito como método mais factível entre as técnicas de avaliação, dada sua maior simplicidade de administração e maior compreensão das tarefas por parte dos respondentes,

apresentando altas taxas de completude, o método apresenta algumas dificuldades. A primeira questão refere-se à inferioridade teórica da métrica em relação às abordagens baseadas na troca. A fim de se converter os valores obtidos a partir da EAV em utilidades SG/TTO, alguns autores procuram estabelecer uma relação sistemática de aproximação através de uma função poder. Entretanto, os coeficientes das funções de transformação diferem sistematicamente entre os estudos (STIGGELBOUT *et al.*, 1996; KRABBE *et al.*, 1997). Enquanto Torrance, Feeny e Furlong (2001) advogam um limitado, mas importante papel da EAV em avaliações econômicas, McCabe e Stevens (2004) criticam esta visão, ao verificarem que a função poder não se mostra robusta em diferentes amostras. Os autores sugerem ainda que esta instabilidade também pode ocorrer entre estados de saúde de uma mesma base de dados.

Além de não representar uma escolha baseada na teoria da decisão, um problema na utilização dos valores da EAV está associado ao efeito de contexto. Segundo Bleichrodt e Johannesson (1997), o valor de um determinado estado na escala está diretamente associado ao número de estados de saúde classificados como melhores ou piores que são apresentados ao indivíduo durante o processo de avaliação. Se um determinado estado aparece diante de estados de saúde melhores, do ponto de vista da qualidade de vida, então seu valor pode ser depreciado. Por outro lado, se o estado de saúde é apresentado em meio a estados mais precários, então tem seu valor elevado. Talvez este efeito possa ser minimizado com a adoção de grupos balanceados que mantêm uma proporção igualitária de estados de saúde de diferentes categorias de severidade.

A literatura distingue os conceitos de preferência, utilidade e valor. Preferência é um conceito mais amplo, que compreende as medidas de utilidade e valor, e que expressa os gostos dos indivíduos. Utilidade refere-se ao resultado de uma decisão sob incerteza, de acordo com o paradigma fundado por von Neumann-Morgenstern (1944), e valor corresponde à preferência medida sob a condição de certeza, de modo a não expressar a atitude de risco (CAMPOLINA e CICONELLI, 2006). Entretanto, em um senso menos estrito da definição de utilidade, e que talvez seja mais relevante para a discussão das medidas de qualidade de vida relacionada à saúde, os termos podem ser tratados como sinônimos na medida que representam um objeto comum – um índice cardinal resumo das preferências por saúde (DOLAN, 2000; TORRANCE, 1986).

Neste sentido, todos os instrumentos baseados nas preferências se propõem a mensurar a satisfação dos indivíduos associada a programas ou estados de saúde. Uma forma de se avaliar diferentes métodos é verificar sua validade, ou seja, analisar em que extensão os instrumentos mensuram o que se propõem. Idealmente, uma métrica poderia ser comparada contra algum critério. Contudo, a falta de um padrão ouro em medidas de preferência em saúde implica no estabelecimento de formas indiretas de se verificar a validade. Uma delas refere-se ao exame da extensão em que os resultados de diferentes métricas convergem entre si. Diversos estudos verificam em que medida os métodos de avaliação se equivalem.

Investigando 58 pacientes com doença renal crônica, Hornberger, Redelmeier e Peterson (1992) determinam a equivalência entre diferentes métodos de obtenção de preferências e encontram baixas correlações e valores médios significativamente diferentes. Os autores atentam para a dificuldade de se escolher o método para a análise de custo-utilidade, dado que esta escolha pode produzir diferenças substanciais nos resultados, com reflexos diretos sobre as políticas públicas. Torrance (1976) encontra uma correlação de Pearson entre TTO e SG igual a 0.65, que o autor interpretou como satisfatória, em contraste com a baixa correlação de 0.36 entre os pesos do SG e da EAV. Read et al (1984) verificam maiores utilidades para o método *Standard Gamble*, quando comparadas àquelas geradas pelo Time Trade-Off, que por sua vez, apresenta valores superiores aos encontrados pela Escala Analógica Visual.

Existem evidências de uma moderada correlação entre os valores obtidos pela EAV e pelos métodos baseados na troca. Comparando cinco métodos de avaliação de estados de saúde, Krabbe, Essink-Bot e Bonsel (1997) analisam a equivalência e a confiabilidade das métricas e observam altos coeficientes de correlação entre todos os métodos analisados, com a escala apresentando valores inferiores aos pesos do TTO e do SG. Além disso, os pesos EAV estão compreendidos em um menor intervalo do potencial espaço de avaliação, sugerindo uma melhor discriminação entre estados de saúde por parte do método da Troca de Tempo. Badia, Monserrat e Herdman (1999), comparando os valores da escala e do TTO, encontram alta correlação entre os métodos, apesar de existirem diferenças na ordenação dos estados de saúde e no número de estados classificados como piores que a morte. Os valores do TTO são maiores que aqueles encontrados pela EAV para estados de saúde brandos e moderados, mas consideravelmente menores para estados mais severos. Dolan *et al.* (1996b) reporta ordenamento similar dos estados de saúde derivados a partir do TTO e do SG. A EAV

apresentou-se como o método mais barato, com menor custo de administração e rapidez em sua realização, mas não necessariamente, um método fácil.

Segundo Robison, Loomes e Jones-Lee (2001), a pronunciada diferença entre os pesos da escala e dos métodos *choice-based* é um reflexo da diferença entre os pontos de referência adotados pelo respondente em cada um dos métodos (KRABBE, ESSINK-BOT e BONSEL, 1997). Na EAV, o indivíduo parece tomar sua própria condição atual de saúde como referência para avaliar os demais estados de saúde. Assim, qualquer problema associado às dimensões de saúde é percebido como uma perda. Nos métodos TTO e SG, a morte corresponde a um dos pontos de referência. Em uma das possibilidades oferecidas nas alternativas, existe a chance de morte. Neste caso, problemas na saúde são percebidos como ganho. Assim, os processos cognitivos que os respondentes utilizam na realização das tarefas são qualitativamente distintos.

Além disso, por envolver probabilidades, o método SG introduz um elemento aleatório adicional na avaliação, tal que os valores obtidos refletem não somente as preferências pelo estado de saúde, mas também o grau de aversão ao risco do indivíduo. Kahneman e Tversky (1979) argumentam que os respondentes geralmente agem com aversão ao risco quando as escolhas são estruturadas em termos de ganhos potenciais, e com propensão ao risco quando as escolhas referem-se à perdas potenciais. O prazer de um ganho possui menor intensidade que a dor de uma perda de tamanho similar. Assim, um estado de saúde deve ser muito indesejável para que o indivíduo aceite um tratamento com risco de morte. Caso contrário, prefere permanecer na condição crônica de saúde com probabilidade unitária. (CRUZ, 2010). Esta atitude mais conservadora diante do risco pode fazer com que as utilidades geradas pelo *Standard Gamble* sejam superestimadas quando comparadas aos valores derivados de técnicas que não envolvem risco. Loomes e McKenzie (1989) mostram evidências de que os indivíduos exibem uma mistura de atitudes de risco e, portanto, suas escolhas não podem ser representadas por um valor constante.

### **3.2 A abordagem ordinal como métrica alternativa e a convergência entre os pesos populacionais**

A métrica ordinal estabelece a ordenação de preferências de duas ou mais alternativas, sem a atribuição do nível destas preferências. Os métodos comumente utilizados incluem o

Experimento de Escolha Discreta (do inglês *Discrete Choice Experiments* – DCE) e o Exercício de Ordenação. Enquanto o primeiro se destaca na avaliação de serviços, produtos e intervenções em saúde, o segundo tem sido aplicado recentemente em avaliações de estados de saúde (BEKKER-GROB, RYAN e GERARD, 2012). No Experimento de Escolha Discreta, os indivíduos escolhem entre duas ou mais alternativas discretas, nas quais ao menos um atributo é alterado. O indivíduo avalia a importância relativa dos estados de saúde de forma a estabelecer uma troca entre os atributos a fim de maximizar sua utilidade associada ao estado.

O Exercício de Ordenação resume-se em uma ordenação de estados de saúde na ordem da preferência do indivíduo. Normalmente, ao indivíduo é solicitado que ordene, de forma descendente, um grupo pré-determinado de estados de saúde. No contexto do protocolo MVH, o exercício de ordenação foi originalmente incluído como um antecessor para métodos cardinais de obtenção de preferência. Assim, o exercício permitiria que os respondentes se familiarizassem com os estados de saúde apresentados, de forma a prepará-los para a avaliação posterior. Apesar das medidas cardinais possuírem âncora numa escala intervalar, a abordagem ordinal possuiria a vantagem da facilidade de compreensão e administração dos exercícios, apresentando um menor erro de medida correspondente.

Dessa forma, diferentemente do que ocorre nos métodos baseados na troca, no Exercício de Ordenação, os indivíduos devem apenas entender o que lhes foi solicitado e refletir sobre a descrição dos estados de saúde e a posição relativa que estes assumem. Assim, esta abordagem apresenta-se como uma forma mais simples de se obter preferências, demandando pouco esforço em termos de suposição de hipóteses, nível de abstração e custo de pesquisa (TORRANCE, 1986). A não necessidade de se introduzir um procedimento de avaliação de forma distinta para os estados classificados como piores que a morte ou de se estabelecer hipóteses sobre a forma da utilidade individual também é considerada um atributo interessante dos exercícios de ordenação em detrimento aos métodos *choice-based*. Assim, a abordagem ordinal supera alguns problemas associados à abordagem cardinal e produz respostas mais consistentes, refletindo as preferências individuais de forma mais precisa (RATCLIFFE *et al.*, 2009).

Nas populações em que se verifica baixo nível educacional, o uso das medidas ordinais seria uma estratégia viável para a aferição das preferências. Muitas populações emergentes,

principalmente países do leste asiático e da América Latina, poderiam se beneficiar do uso das técnicas ordinais para guiar as políticas públicas nesta área. Ademais, a incorporação das preferências da população juvenil e idosa e a diminuição de exclusões de dados inconsistentes trariam grandes avanços em termos de representatividade amostral (BANSBACK *et al.*, 2012b). Coaste *et al.* (2008) aponta a complexidade das tarefas de métricas como TTO e SG para a população idosa. Torna-se menos complexo prover respostas sobre a preferência relativa entre dois estados de saúde do que quantificar a magnitude destas preferências em diferentes cenários hipotéticos. Além disso, métodos cardinais podem causar confusão mental nos respondentes, particularmente, quando se faz referência à morte.

Uma outra questão se refere ao alto custo das pesquisas envolvendo técnicas *choice-based*, como TTO e SG (STIGGELBOUT *et al.*, 1996). A logística de campo empreendida torna-se complexa, na medida em que o treinamento dos pesquisadores e o processo de entrevistas podem se tornar morosos, quando comparada à logística necessária para a aplicação da ordenação, por exemplo. Neste sentido, o uso de abordagens relativamente mais simples pode representar uma alternativa interessante do ponto de vista dos critérios de razoabilidade das pesquisas científicas, tornando “custo-efetivas” as próprias análises custo-utilidade. Contudo, as informações obtidas a partir desta abordagem não representam utilidades Von Neumann-Morgenstern, pois não requerem que os respondentes realizem trocas, mas somente considerem a satisfação relativa entre estados de saúde. Além disso, é necessário ter cautela no uso de informações ordinais para a obtenção dos pesos populacionais, na medida que a utilização destas informações em análises custo-utilidade requer o estabelecimento de âncoras para os valores estimados a partir de modelos estatísticos (BRAZIER *et al.*, 2012).

Recentemente, dados na forma ordinal têm sido usados para a derivação de valores cardinais de estados de saúde, a partir de modelos logísticos, utilizando uma grande variedade de sistemas descritivos. Analisando dados para o Reino Unido a partir do sistema EQ-5D, Salomon (2003) encontra grande pareamento entre os valores estimados na ordenação e no método da Troca de Tempo, reforçando a inclusão do exercício de ordenação em outros *surveys*, caso ainda não foram incorporados nos protocolos. Resultados similares foram encontrados por Craig, Busschbach e Salomon (2009a), examinando dados de oito países: Eslovênia, Argentina, Dinamarca, Japão, Holanda, Espanha, Reino Unido e Estados Unidos.

Utilizando o modelo logit condicional nos dados ordenados, os autores observam uma grande distância entre os valores para o estado de saúde perfeita e os valores do próximo melhor estado. Além disso, os valores baseados no método ordinal apresentaram maior diferenciação entre si, quando comparados com os valores derivados a partir das métricas TTO e EAV, sugerindo algumas contradições dos resultados com o construto teórico dos exercícios cardinais. Os autores também sugerem que o uso do estado de saúde 33333 como âncora é mais apropriado que o uso do estado ‘morte’, devido à heterogeneidade de crenças religiosas com respeito à morte. O método de ordenação gera menor proporção de estados de saúde piores que a morte que o método cardinal.

McCabe *et al.* (2006) estimam modelos de regressão logístico condicional utilizando dados ordinais do *Health Utilities Index Mark 2* e SF-6D para a população do Reino Unido. Diferentemente de Salomon (2003), que reescala os coeficientes do modelo utilizando a média do TTO referente ao estado de saúde 33333, os autores evitam o uso de dados externos, pois estimam diretamente um parâmetro para o estado Morte. Quando comparados aos modelos de regressão convencionais, estimados a partir do método cardinal *Standard Gamble*, verifica-se grande proximidade da performance preditiva. Enquanto o SG apresentou uma leve vantagem em termos preditivos, o número de inconsistências foi menor no exercício de ordenação.

Utilizando dois sistemas descritivos de saúde específicos, Brazier *et al.* (2012) estima utilidades para estados de saúde a partir de informações ordinais. Os coeficientes encontrados a partir do Exercício de Ordenação são muito similares aos encontrados pelo método TTO, com comportamento esperado quanto aos sinais e a magnitude dos efeitos associados aos níveis de severidade de cada dimensão. Os erros absolutos médios apresentaram-se inferiores a 0.10 para ambos os modelos – 0.056 e 0.061 para o TTO e a ordenação, respectivamente. O número de predições com erro absoluto superior a 0.05 foi maior com o modelo para dados ordinais e a correlação com os pesos observados no TTO foi igual a 0.715. A modelagem com os dados obtidos a partir do DCE apresenta pior ajuste na comparação com o método da Troca de Tempo. Alguns coeficientes estimados apresentam sinais contra-intuitivos, além do maior erro absoluto médio, apesar de permanecer inferior a 10%. Utilizando DCE para estimar preferências de pessoas idosas, Ryan *et al.* (2006) encontrou sinais e magnitudes esperados para os coeficientes estimados.

Rand-Hendriksen et al (2012), utilizando dados do Reino Unido e Estados Unidos, analisam se os métodos de ordenação e TTO diferem no que se refere à sensibilidade das dimensões do sistema descritivo EQ-5D. Diferenças de médias positivas, ou seja, estados de saúde são classificados com uma pontuação menor no TTO que na ordenação, são comuns nos estados de saúde dominados por problemas que envolvem dor/desconforto e ansiedade/depressão. Diferenças de médias negativas, refletindo estados de saúde sendo classificados com pontuação menor na ordenação que no TTO, são associadas às dimensões de mobilidade, autocuidado e atividades usuais. Os autores alertam para a questão de qual método melhor representa as preferências da população e às dificuldades que este resultado traz para as avaliações de políticas públicas.

### **3.3 Inconsistência Interna: o trade-off entre a qualidade das informações e a representatividade amostral**

A análise das vantagens e desvantagens de técnicas de avaliação em saúde deve levar em consideração não somente a extensão que um instrumento mede o que se propõe a mensurar, mas principalmente, a capacidade de se produzir respostas consistentes de acordo com a racionalidade proposta pelo estudo. Dado que as avaliações de estados de saúde são baseadas em cenários hipotéticos, a saúde do respondente não é incorporada na decisão de forma direta. Assim, a qualidade dos dados depende da disposição do indivíduo em participar e revelar suas preferências de forma honesta e com suficiente nível de esforço (CRAIG e RAMACHANDRAN, 2006). Além disso, requer que o objeto em análise seja passível de classificação em uma escala de ordenação; que o participante tenha alguma capacidade cognitiva para realizar as tarefas, e perceba a relação entre suas decisões e o objetivo do estudo; e por fim, que seu incentivo em participar seja suficiente, de forma a superar os incentivos à sabotagem. Se alguma dessas premissas for violada, o desenho do estudo pode enviesar os resultados, com conseqüências para a formulação de políticas públicas direcionadas à comunidade de um modo geral.

Não existe um consenso relacionado ao conceito de inconsistência, sendo este, definido por cada autor, dependendo do contexto do estudo. Neste sentido, diversos tipos de inconsistências são apresentados na literatura: i) indivíduos que avaliaram menos que três estados de saúde; ii) indivíduos que atribuem um único valor a distintos estados de saúde; iii) indivíduos que não avaliam os estados âncoras, Saúde Perfeita e Morte; iv) indivíduos que

atribuem à Morte, valor maior ou igual àquele atribuído ao estado de Saúde Perfeita, e de forma mais abrangente; (v) indivíduos que atribuem a um estado de saúde pior, valor **maior** àqueles atribuídos a estados de saúde melhores, em termos de qualidade de vida - monotonicidade estrita das preferências (CRAIG e RAMACHANDRAN, 2006; DOLAN e KIND, 1996; BADIA, MONSERRAT e HERDMAN, 1999; DELVIN *et al.*, 2003; KRABBE *et al.*, 1997; DEVLIN, HANSEN e SELAI, 2004).

Diante destas escolhas ditas irracionais, a censura de respostas inconsistentes da análise final dos resultados tem sido o tratamento empregado na literatura de estimação de parâmetros populacionais de estados de saúde. No estudo de obtenção de pesos para a população canadense, por exemplo, dos 2394 respondentes que aceitaram participar da pesquisa, apenas 1145 (48%) são incluídos na análise primária. Foram excluídos os indivíduos que não completaram todas as avaliações, que atribuíram mesmo valor a todos os estados de saúde e que obtiveram mais que um par de estados de saúde com avaliações reversas à ordem lógica esperada (BANSBACK *et al.*, 2012a). Sackett e Torrance (1978) consideram apenas as informações de indivíduos que avaliaram todos os cenários propostos no exercício e que foram consistentes quanto à ordem lógica pré-estabelecida pelo instrumento, eliminando 22% dos indivíduos da amostra.

No estudo de estimação dos parâmetros dos estados de saúde definidos a partir do EQ-5D, para a população de Taiwan, Lee *et al.* (2013) excluem 38,79% dos respondentes da amostra. Estes participantes cometeram pelo menos um dos seguintes tipos de inconsistências: todos os estados de saúde são classificados como piores que a morte, atribuição de um mesmo valor para todos os estados de saúde e, possui quatro ou mais inconsistências lógicas, ou seja, violação da ordem lógica presente no instrumento EQ-5D. Os respondentes excluídos são, em sua maioria, idosos, com menor nível de escolaridade e que, em média, levaram maior tempo para completar a entrevista.

Segundo Kind e Dolan (1996), as inconsistências distinguem-se em primárias e secundárias. Enquanto as primeiras estão ligadas às limitações intrínsecas aos respondentes, as segundas referem-se às conseqüências de algum aspecto do procedimento de medida. A questão envolvida na inconsistência primária é a frequência com que estas ocorrem em subgrupos populacionais específicos e a exclusão de respostas inconsistentes da análise final dos resultados. A remoção dos respondentes logicamente inconsistentes do banco de dados, altera

de forma significativa a representatividade amostral e, em último caso, os AVAQS ganhos em avaliações de tecnologias em saúde. Se por um lado, a arbitragem na remoção de dados inconsistentes pode sub-representar as preferências de grupos específicos da população, por outro lado, a inclusão dos respondentes que, a priori, oferecem informações consideradas contrárias à racionalidade esperada, pode produzir estimativas enviesadas dos pesos sociais dos estados de saúde. Assim, a não ser que se invista em melhorias dos instrumentos de coleta de dados e no aperfeiçoamento destas técnicas a contextos sociais e demográficos distintos, existirá um *trade-off* entre qualidade e representatividade dos dados amostrais (DEVLIN *et al.*, 2003).

Isto se torna particularmente importante se as preferências destes indivíduos excluídos diferem dos gostos de outros grupos populacionais ou da média geral. Como algumas características sociodemográficas e de saúde têm mostrado possuir efeitos significativos sobre as preferências por estados de saúde (DOLAN *et al.*, 1996a; KIND *et al.*, 1998; LUNDENBERG *et al.*, 1999), parecem essenciais as tentativas de se melhorar a consistência por parte dos indivíduos idosos, com menores níveis de educação formal e de condição socioeconômica mais precária. No que se refere às inconsistências secundárias, talvez não seja surpreendente que os métodos baseados no *trade-off*, também chamados de métodos *choice-based*, produzam maior número de inconsistências que os demais, na medida em que o Exercício de Ordenação e a Escala Analógica Visual possuem a vantagem de os respondentes visualizarem todos os estados de saúde de forma simultânea, além de requererem menor grau de abstração no seu desenvolvimento.

Verificando os dados de Barcelona a partir de três métodos de obtenção de pesos, Badia *et al.* (1999) encontram percentuais de indivíduos inconsistentes iguais a 59.2%, 25.9% e 24.4%, para as métricas da Troca de Tempo, EAV e Ordenação, respectivamente. Quando se considera a razão entre os pares de estados de saúde inconsistentes e os pares potencialmente inconsistentes, o TTO também apresenta maior taxa média e mediana de violação da ordem lógica do EQ-5D. Enquanto a taxa média de inconsistências registra valor de quase 4% no método da Troca de Tempo, as taxas verificadas no Exercício de Ordenação e na Escala Analógica Visual são de 1.32% e 1.14%, respectivamente. Além disso, aumento da idade e baixos níveis de escolaridade são positivamente associados com altos níveis de inconsistência nos métodos cardinais, apesar de não haver qualquer efeito quando se considera o Exercício

de Ordenação. Krabbe, Essink-Bot e Bonsel (1997) encontram uma taxa de inconsistência média de 4,6%, 4,3% e 2% para os métodos SG, TTO e EAV, respectivamente.

Utilizando dados do Reino Unido e Estados Unidos, Craig e Ramachandran (2006) verificam o efeito das características dos indivíduos sobre as inconsistências lógicas, e o efeito destas sobre os pesos populacionais para estados de saúde. Os autores encontram que respondentes inconsistentes são, em grande parte, idosos, mulheres, de níveis educacionais mais baixos, que apresentam dificuldades em executar as tarefas do TTO e da EAV e apresentam pior saúde auto-avaliada segundo o sistema descritivo EQ-5D. A remoção dos respondentes inconsistentes modificou os valores médios do TTO e da EAV para diversos estados de saúde. Estados brandos tiveram seus valores aumentados, enquanto estados severos apresentaram queda dos seus valores, aumentando a dispersão dos pesos médios. Bravata *et al.* (2005) encontra uma relação direta entre o tempo da entrevista e o nível de inconsistências.

No estudo para Malásia, que investigou 152 pacientes de nove hospitais públicos, o método TTO foi considerado mais difícil que a EAV, além de apresentar maior número de inconsistências lógicas. Considerando o conceito de monotonicidade estrita das preferências, enquanto cerca de 63% dos respondentes foram totalmente consistentes na Escala Analógica Visual, apenas 15.7% deles assim o foram no método da Troca de Tempo (YUSOF, GOH e AZMI, 2012). Utilizando informações da EAV, no estudo para Nova Zelândia, Devlin *et al.* (2003) encontram um percentual de 21% de respondentes sem qualquer inconsistência em suas avaliações. O grupo de indivíduos que apresentaram, no máximo, uma inconsistência, formou a subamostra de maior proximidade com a amostra completa. Os coeficientes estimados para cada nível e dimensão de saúde mantiveram a importância relativa nas duas amostras e os pesos estimados não apresentaram diferenças significativas.

Lamers *et al.* (2006), utilizando as avaliações dos estados de saúde definidos pelo EQ-5D de 309 indivíduos adultos da Holanda, encontram proporções iguais a 65% e 89% de respondentes inconsistentes na EAV e no TTO, respectivamente. Os valores observados na Escala Analógica Visual tendem a ser menores para os indivíduos consistentes, enquanto que no método da Troca de Tempo, pesos mais baixos estão associados aos respondentes inconsistentes. A remoção dos respondentes que apresentaram mais que três inconsistências, não resultaram em diferenças significativas dos coeficientes e pesos estimados na amostra completa. A partir de avaliações realizadas com a EAV, Dolan e Kind (1996) encontram

maiores inconsistências associadas a indivíduos idosos e com menor nível educacional. O sexo do respondente não está correlacionado com o grau de consistência.

A ocorrência de inconsistências de um modo geral e, de forma especial, quando os indivíduos atribuem a um estado de saúde pior, valor **maior ou igual** àqueles atribuídos a estados de saúde melhores (monotonicidade fraca ou não-estrita das preferências), capta diversos tipos de comportamento. A tarefa de avaliar estados de saúde pode ser cognitivamente difícil. A complexidade e a morosidade de determinados métodos podem imprimir cansaço ou confusão mental aos indivíduos, principalmente aos idosos e àqueles com menor nível de escolaridade. Educação formal pode indicar maior tolerância à tarefas árduas, do ponto de vista cognitivo, e maior motivação a suportar pesquisas científicas (grau de conscientização). Dessa forma, alguns indivíduos podem falhar em entender as instruções, em compreender a natureza hipotética dos estados de saúde e ter dificuldades com o conceito de morte, contribuindo para o aumento de exercícios sem avaliações, questionários com apenas um ou dois estados avaliados ou com todos os estados apresentando o mesmo peso ou valor (CRAIG e RAMACHANDRAN, 2006).

Respondentes *non-traders* são aqueles que nunca estão dispostos a trocar uma redução no tempo de vida a fim de se melhorar a qualidade de vida relacionada à saúde (ROBSON, DOLAN e WILLIAMS, 1997). Neste sentido, a escolha em relação à morte é determinística, contrariando as hipóteses dos modelos *choice-based*. Esse comportamento talvez possa ser explicado por crenças religiosas e espirituais, na medida em que diversos respondentes afirmam que nenhum estado de saúde pode ser pior que a morte, pois a experiência naquela condição de vida representa um desígnio de Deus e não está sujeita a escolha humana. Além disso, alguns indivíduos podem interpretar a preferência pela morte em detrimento a um estado de saúde, como a interrupção voluntária da vida através de eutanásia ou suicídio. Segundo Devlin, Hansen e Selai (2003), que apresentam uma análise qualitativa dos comentários dos respondentes, outro comportamento inconsistente relacionado às crenças religiosas ocorre quando o valor atribuído à Morte é superior ou igual àquele atribuído à Saúde Perfeita. Nestes casos, alguns indivíduos, principalmente cristãos, também consideram que após a morte, não possuirão qualquer problema em nenhuma dimensão.

Respondentes sabotadores são aqueles que se rebelam contra o objetivo do estudo a fim de abortar o procedimento o mais rapidamente possível, seja por desinteresse ou por cansaço. No

início da entrevista, os indivíduos podem se mostrar dispostos a realizar as tarefas. Contudo, com o decorrer do processo, sentem-se desmotivados devido à morosidade da entrevista. Neste sentido, a ordem comumente adotada de realização dos instrumentos, que apresenta nas fases iniciais os protocolos de Ordenação e Escala Analógica Visual, pode imprimir diferenças na consistência dos indivíduos ao realizarem as avaliações. Os indivíduos também podem não finalizar as tarefas e atribuir valores iguais a diferentes estados de saúde, de modo a atenuar as diferenças das tarifas sociais, não devido às preferências individuais particulares, do ponto de vista modal das escolhas, mas graças ao comportamento no momento da entrevista (CRAIG e RAMACHANDRAN, 2006).

O entrevistador também pode impor forte influência na geração de inconsistências. Segundo Badia, Monserrat e Herdman (1999), para que a pesquisa resulte em informações de qualidade, os entrevistadores necessitam não somente de treinamento apropriado, mas também, de prévia experiência em trabalhos que envolvem pesquisa de campo e condução de questionários. Os resultados encontrados revelam que a taxa de inconsistência em instrumentos conduzidos por entrevistadores menos experientes foi o dobro daquela encontrada em questionários conduzidos pelos pesquisadores mais ambientados a este tipo de tarefa.

A dificuldade intrínseca à conceitualização das inconsistências surge da complexidade de se julgar como inconsistentes, os comportamentos ditos irracionais - do ponto de vista lógico do exercício praticado - quando, de fato, estes podem estar refletindo as preferências individuais. Uma evidência que ilustra este fato são os respondentes que experimentaram alguma condição de saúde ou trauma social e incorporaram a adaptação em suas avaliações, atenuando as diferenças percebidas entre estados de saúde (GROOT, 2000). Nesse contexto, pessoas idosas e menos saudáveis possuem maior experiência com situações precárias de saúde e podem, via processo de adaptação, reverterem a ordem lógica esperada ou atribuírem valores semelhantes a estados de saúde marcadamente distintos no que se refere ao grau de severidade de seus domínios.

Este fenômeno também é retratado na literatura, em análises de decomposição dos fatores determinantes da felicidade. Observa-se que alguns eventos transitórios ou inesperados, como ganhar na loteria ou sofrer um acidente grave, possuem efeitos significativos no momento de ocorrência, alterando os níveis de felicidade. Contudo, com o decorrer do tempo, os índices de

bem-estar tendem a retornar aos seus níveis originais (BRICKMAN, COATES e JANOFF-BULMAN, 1978). Assim, o fenômeno da adaptação tem sido frequentemente utilizado para explicar a baixa correlação entre felicidade e renda. Neste contexto, a tese atacaria o pressuposto microeconômico de monotonicidade, impondo um sentido de saciedade, a partir do qual o progresso material não afetaria o bem-estar (FRANCO, 2012).

Como destacado por Craig e Ramachandran (2006), a utilização de ferramentas computacionais que auxiliam os entrevistadores na detecção de respostas inconsistentes, a fim de que estes alertem os participantes a revisarem suas escolhas, pode melhorar as estimativas em detrimento do aumento dos custos financeiros e do tempo médio de pesquisa, desfavorecendo a logística de campo. Ademais, avanços no controle de respostas inconsistentes podem resultar em informações que não refletem as decisões do respondente e, conseqüentemente, os gostos sociais. Assim, o uso de abordagens mais simples na obtenção das preferências em saúde, como o Exercício de Ordenação, pode configurar-se como uma alternativa interessante.

## 4. ASPECTOS METODOLÓGICOS GERAIS

Esta seção é composta por três subseções. Na primeira, apresenta-se a descrição da base de dados e dos protocolos utilizados na avaliação dos estados de saúde. Na segunda subseção, são descritos o modelo econométrico utilizado na transformação das informações ordinais em resultados cardinais e os procedimentos de reescalonamento destes valores estimados a fim de se adotar as âncoras exigidas para análises de custo-utilidade. Ainda nesta subseção são apresentadas as medidas de comparação utilizadas para a análise de equivalência das utilidades derivadas de diferentes métodos. Na terceira e última subseção, apresentam-se o conceito e o cálculo das inconsistências lógicas.

### 4.1 Base de Dados

A base de dados utilizada neste trabalho foi obtida a partir dos resultados de uma pesquisa inédita realizada pelo Centro de Desenvolvimento e Planejamento Regional - CEDEPLAR/UFMG e financiada pela FAPEMIG - Fundação de Amparo à Pesquisa do estado de Minas Gerais. Inserido no edital Programa Pesquisa Para o SUS - PPSUS, o projeto “Estimação dos Parâmetros de Valorização dos Estados de Saúde em Minas Gerais a partir do EQ-5D”, coordenado pelas professoras Mônica Viegas e Kenya Noronha, foi conduzido no estado de Minas Gerais, em 2011. Os primeiros resultados das estimações dos pesos para o Brasil já foram publicados e diferenças nas distribuições dos parâmetros entre os países, inclusive entre os latinos americanos, foram identificadas, reforçando o peso que a importação de tecnologia de avaliação possui sobre os resultados, e a necessidade da utilização de valores do próprio país (ANDRADE *et al.*, 2013; GALANTE *et al.*, 2011).

Anterior ao estudo em questão, pesos para o país foram obtidos a partir de uma pesquisa de campo realizada na cidade de Porto Alegre (RS) e liderada pelo Instituto de Avaliação de Tecnologias em Saúde – IATS, da Universidade Federal do Rio Grande do Sul - UFRGS. Contudo, a extrapolação dos resultados derivados desta pesquisa requer ainda mais cautela, dada as particularidades socioeconômicas do estado do Rio Grande do Sul em relação ao restante do país. Além disso, a análise utilizou sistema descritivo e métodos de obtenção de preferências distintos aos adotados em Minas Gerais, com a adoção do SF-6D e da métrica *Standard Gamble* (CRUZ *et al.*, 2011).

Para realização da pesquisa de campo foram investigados indivíduos alfabetizados, com idade entre 18 e 64 anos, residentes de áreas urbanas de Minas Gerais. A definição do tamanho amostral foi baseada no Censo Demográfico de 2010, de modo a obter uma margem de erro estimada em 3%. No total, 3362 indivíduos foram recrutados, dos quais 1115 viviam na capital Belo Horizonte, 626 residiam na Região Metropolitana de Belo Horizonte e 1621 no interior do Estado. Considerou-se a técnica de amostragem probabilística, estratificada por sexo e idade, para todo o estado e para os três distintos níveis regionais. Além disso, a amostra foi espacialmente distribuída de forma a considerar todas as macrorregiões de Minas Gerais e todas as zonas de planejamento de Belo Horizonte.

As características demográficas e socioeconômicas da amostra são bastante comparáveis aos dados fornecidos por outros bancos de dados para o estado de Minas Gerais, como a Pesquisa por Amostra de Domicílios - PAD 2009 e a Pesquisa Nacional por Amostra de Domicílios – PNAD 2008. Segundo a TABELA 2, a distribuição por idade e sexo possui forte similaridade, indicando a eficácia da amostra em atingir representatividade destas variáveis para a população alfabetizada do estado. O intervalo etário pesquisado foi entre 18 e 64 anos, sendo a idade média do respondente igual a 38 anos. As entrevistas duraram, em média, 44 minutos. Dos 3362 indivíduos recrutados, 177 avaliaram menos que sete cartões no Exercício de Troca de Tempo e 57% reportaram dificuldade com este instrumento (ANDRADE *et al.*, 2013).

O estado de Minas Gerais apresenta grande heterogeneidade em termos de desenvolvimento econômico e padrão de vida. A análise do Índice de Desenvolvimento Humano (IDH) mostra evidências de quão similares são as desigualdades socioeconômicas presentes no estado se comparados àquelas observadas para o país. Em 2000, os valores do IDH para Minas Gerais variaram dentro do intervalo de 0.57 (região nordeste do estado) a 0.84 (região sudeste do estado), enquanto no Brasil, este intervalo foi de 0.64 (Maranhão) a 0.82 (Santa Catarina) (PNUD, 2003). Assim, a grande diversidade dentro do próprio estado pode ser considerada representativa da heterogeneidade da população brasileira, viabilizando, mesmo que com cautela, possíveis extrapolações do estudo.

O trabalho de campo confirmou as evidências internacionais com respeito à complexa logística de aplicação dos instrumentos cardinais, a nível populacional. Uma agência de pesquisa de mercado foi contratada para a concretização desta etapa do estudo. Foram realizados diversos encontros entre a equipe contratada e os participantes do Grupo de

Estudos em Economia da Saúde e Criminalidade - GEESC/UFMG, a fim de se discutir a adequação dos instrumentos de coleta de dados quanto aos seus aspectos operacionais, além do planejamento de todas as atividades subjacentes e do cronograma para as etapas do projeto. Um pré-teste foi conduzido por dois pesquisadores da agencia de pesquisa, que aplicaram quatorze entrevistas durante dois dias de trabalho. Estimou-se uma produção diária de cerca de 3-4 entrevistas por pesquisador, com duração média de 45 minutos para cada questionário. Após tal procedimento, os instrumentos sofreram algumas modificações e foram aprovados de forma definitiva.

**TABELA 2 – Características sociodemográficas da amostra e comparação com bases de dados nacionais (%)**

<b>Características</b>	<b>PPSUS</b>	<b>FJP</b>	<b>IBGE</b>
<b>Sexo</b>			
Homem	51,58	52,43	52,08
Mulher	48,42	47,57	47,92
<b>Faixa Etária</b>			
18-34	43,3	47,23	46,2
35-49	33,95	32,5	33,37
50-59	16,25	15,29	15,9
>60	6,5	4,98	5,13
<b>Nível Educacional</b>			
Primário Incompleto	4,86	-	5,62
Primário Completo	24,37	-	7,7
Fundamental Completo	24,56	-	37,51
Médio Completo	37,64	-	36,43
Superior Completo	8,54	-	12,74
<b>Estado Civil</b>			
Casado	56,12	45,85	-
Viúvo	2,83	2,95	-
Divorciado ou Separado	7,92	6,71	-
Solteiro	33,13	44,41	-
<b>Plano de Saúde</b>			
Sim	31,36	28,34	35,38
Não	68,64	71,28	64,62
<b>Saúde auto-avaliada</b>			
Muito bom	25,35	29,17	31,18
Bom	52,01	49,41	48,99
Regular	20,49	18,32	17,12
Ruim	1,58	2,4	2,14
Muito ruim	0,49	0,65	0,57

Fonte: Andrade *et al.*, 2013.

Um total de 25 entrevistadores, dois auxiliares de coordenação e cinco profissionais de revisão e codificação foi recrutado através da agencia de pesquisa de mercado. Com o apoio da equipe GEESC/UFMG, o treinamento dos pesquisadores foi realizado em três dias, com registro audiovisual. Além de participarem das palestras de instrução, os pesquisadores

aplicaram alguns questionários a título de teste. Dada a complexidade do instrumento, dos 25 pesquisadores recrutados, cinco desistiram do projeto e sete foram reprovados ao longo de toda pesquisa de campo, mesmo após novas rodadas de treinamento. Os pesquisadores utilizaram coletes de identificação, que foram confeccionados juntamente com 28 pranchetas ou Quadros de Troca do Tempo, necessários para a aplicação das entrevistas.

A coleta de dados foi realizada entre outubro e dezembro de 2011, apresentando um resultado médio diário de cinco questionários por pesquisador. As entrevistas foram conduzidas de forma a selecionar apenas um indivíduo de cada domicílio e incentivos econômicos não foram fornecidos aos respondentes. Após a entrega dos questionários pelos entrevistadores, uma equipe de dois coordenadores e quatro revisores, supervisionada pelos integrantes do GEESC/UFMG, se concentrou em avaliar a qualidade dos materiais. Cerca de 20% dos questionários foram checados por telefone a fim de detectar possíveis fraudes por parte dos entrevistadores. Ademais, a maioria dos pesquisadores precisou receber novas instruções sobre a aplicação dos instrumentos, de modo que os erros encontrados nos primeiros lotes da pesquisa não voltassem a se repetir. Dessa forma, procurou-se garantir a qualidade na coleta dos dados e a minimização do erro sistemático dos pesquisadores na aplicação dos protocolos.

A versão brasileira do sistema descritivo EQ-5D foi culturalmente adaptada e o protocolo de entrevistas seguiu uma versão revisada (KIND, 2009) do estudo original do MVH (GUDEX, 1994). Este novo formato já havia sido aplicado nas avaliações dos pesos populacionais para França e Coréia do Sul (CHEVALIER e POUVOURVILLE, 2011; LEE, *et al.*, 2009). A fim de minimizar os espaços de interpolação, dos 243 estados de saúde teóricos, 102 estados foram obtidos de forma direta, seguindo a métrica de distância das dimensões de cada estado em relação ao índice de saúde perfeita.<sup>5</sup> Os estados foram agrupados em 26 blocos, cada um composto por seis perfis de saúde (dois brandos, dois moderados e dois severos). Devido ao grande tamanho da amostra, mais de 100 observações por estado de saúde foram garantidas, de acordo com as recomendações de Chuang e Kind (2010).

Somado ao bloco de estados de saúde, cada indivíduo também avaliou os cartões definidos logicamente como o melhor e o pior estados de saúde hipotéticos (estados 11111 e 33333, respectivamente), além do cartão 'Morte', totalizando nove estados de saúde. Neste sentido,

---

<sup>5</sup> O procedimento para classificação dos estados de saúde quanto aos níveis de severidade é detalhado na seção 4.2.3. Para mais informações, ver Kind (2009).

foram realizadas importantes inovações metodológicas se comparado às pesquisas conduzidas anteriormente, principalmente, na Europa e Estados Unidos, dado que se investigavam somente 42 estados de saúde de forma direta, interpolando os 200 estados restantes através de análise de regressão. As descrições de cada perfil foram apresentadas ao respondente sob a forma de conjuntos de cartões impressos.

O questionário consiste de três instrumentos. No Instrumento A, ao indivíduo é solicitado que descreva seu próprio estado de saúde, em termos do sistema descritivo EQ-5D, e registre uma pontuação equivalente para a condição descrita através da Escala Analógica Visual, com as âncoras de 0 e 100 correspondendo ao pior e ao melhor estados de saúde imagináveis. Depois, ordena de forma descendente um grupo pré-determinado e balanceado de nove estados de saúde, atribuindo-lhes diferentes posições. Ainda neste mesmo instrumento, realiza-se o preenchimento da Escala Analógica Visual, indicando em que local da escala os participantes classificam os nove cartões. Entre a aplicação do exercício de ordenação e o preenchimento da EAV, os cartões são apresentados de forma aleatória ao entrevistado, a fim de que a evidência de correlação entre os dois métodos seja espontânea. É explicado que cada estado de saúde deve durar por 10 anos sem qualquer mudança, seguido por morte. Estes exercícios são realizados anteriormente ao método da Troca de Tempo com a finalidade de familiarizar os indivíduos às descrições dos perfis de saúde.

No Instrumento B, aos participantes lhe são apresentadas escolhas entre alternativas que variam entre níveis de quantidade e qualidade de vida<sup>6</sup>. Cada indivíduo valoriza sete estados de saúde (seis estados que abarcam os três níveis de severidade e o pior estado de saúde hipotético 33333), utilizando o método de Troca de Tempo através de um quadro de dupla-face<sup>7</sup>, no qual em um dos lados se avaliam estados considerados pelos respondentes como melhores que a morte e, na outra face, estados registrados como piores que a morte. No primeiro caso, o respondente seleciona o número  $x$  de anos ( $x < 10$ ) sob o melhor estado de saúde hipotético (Saúde Perfeita) que equivale a viver por 10 anos no estado com algum decréscimo de saúde; quanto menor este período de equivalência, pior é o estado em questão. No caso dos estados piores que a morte, os indivíduos escolhem morrer imediatamente ou viver por um número  $x$  de anos ( $x < 10$ ) na condição específica de saúde, seguido por  $(10 - x)$  anos no estado de saúde perfeita como forma de compensação. Assim, quanto maior o tempo

---

<sup>6</sup> Os instrumentos A e B são apresentados no Anexo A.

<sup>7</sup> Consulte o Apêndice D para ver a figura do Quadro de Troca de Tempo.

requerido no melhor estado de saúde para compensar um curto período de tempo vivido na condição em questão, pior é este estado (GUDEX, 1994). O Instrumento C considera informações referentes às características demográficas e socioeconômicas de cada respondente.

Se saúde integral e morte apresentam pontuações 1 e 0 respectivamente, então para estados classificados como melhores que a morte no TTO, seus valores são dados pela fórmula

$V = \frac{x}{10}$ , onde x é o período de tempo vivido com saúde integral. Para estados classificados

como pior que a morte, seus valores são dados pela fórmula  $V = \frac{-x}{(10-x)}$ . Contudo, este

procedimento gera valores com limite inferior igual a -19, enviesando para baixo os pesos. A exemplo de DOLAN *et al.* (1996a), a fim de tratar esta distribuição assimétrica dos valores

negativos, realizou-se uma transformação monotônica através da fórmula  $V_t = \frac{V}{(1-V)}$ , se  $V <$

0.

Dessa forma, estados classificados como piores que a morte foram limitados inferiormente pelo valor -1, trabalhando-se com um intervalo equidistante da morte em ambas direções, positiva e negativa. Esta transformação é comumente utilizada na literatura.<sup>8</sup> As pontuações obtidas a partir da Escala Analógica Visual também sofrem transformações, a fim de que os estados de saúde perfeita e morte representassem âncoras para todos os respondentes. Assim,

segundo a fórmula  $EAV_j^t = \frac{EAV_j - EAV_{MORTE}}{EAV_{PERFEITA} - EAV_{MORTE}}$  obtém-se valorações padronizadas, de

modo a fixar os valores das âncoras em 0 (Morte) e 1 (Saúde Perfeita).

#### **4.2 Transformação de informações ordinais em dados cardinais: o modelo logit condicional**

Diante do contexto de ampliação do escopo de instrumentos alternativos para a obtenção de pesos sociais, verifica-se a viabilidade do Exercício de Ordenação em avaliações de tecnologias em saúde. Contudo, um dos desafios presentes na utilização de informações ordinais neste tipo de avaliação econômica refere-se ao caráter qualitativo de seus dados, o

---

<sup>8</sup> Para diferentes transformações dos pesos ver Lamers (2007).

que dificulta a produção de resultados cardinais. Através do modelo logit condicional, esta transformação pode ser realizada.

#### 4.2.1 Exercício de Ordenação

O modelo do Julgamento Comparativo (THURSTONE, 1927) é identificado na literatura psicológica como a base teórica para a derivação cardinal a partir de dados ordinais. A Lei de Thurstone considera que o indivíduo registra diferentes julgamentos comparativos em sucessivas ocasiões sob um mesmo par de estímulos. No entanto, existe um julgamento comparativo para determinado estímulo que é mais comumente experimentado que outros, chamado processo discriminatório modal. A separação entre um julgamento ordinário e o processo modal gera um desvio discriminatório na ocasião. Logo, o processo discriminatório de um dado estímulo não é fixo, mas flutua em torno de um nível modal, ou seja, existe um termo de erro na função de utilidade latente individual.

Uma escala contínua psicológica é definida a partir das frequências destes processos de julgamentos de forma que, quanto mais próximas estiverem as utilidades latentes dos dois estímulos, A e B, maior a probabilidade de um indivíduo dizer que prefere A a B, quando de fato, a utilidade que ele espera ganhar no estímulo A é menor que a utilidade que espera obter em B. A abordagem de Thurstone tem sido constantemente modificada de forma a incluir a aplicação da função logística como meio de modelagem dos valores latentes implícitos em dados ordinais.

A análise dos dados do Exercício de Ordenação é baseada em McFadden (1974), que propôs um modelo de utilidade aleatória como meio de estimar a função de utilidade latente. Este modelo é operacionalizado usando o *conditional logit model*, que também tem sido referido como *rank-ordered logit* ou *exploded logit model* (SALOMON, 2003; McCABE *et al.*, 2006; CRAIG *et al.*, 2009, BRAZIER *et al.*, 2009; RYAN *et al.*, 2006). Assume-se que em exercícios de ordenação, o respondente realiza uma série de seleções num espaço de conjuntos de estados de saúde cada vez menores.

Então, em uma ordenação de  $n$  cartões, primeiramente cada indivíduo escolhe o estado de saúde preferível; depois, escolhe aquele preferível dentre os  $n-1$  estados de saúde restantes; e assim sucessivamente, até que todos os estados de saúde estejam ordenados de 1 a  $n$ . Cada

respondente ordena  $J$  estados de saúde sendo  $Y_{ij}$  a posição dada para o estado  $j$  pelo indivíduo  $i$ . Assume-se que o respondente  $i$  possui um valor de utilidade latente para o estado  $j$ ,  $U_{ij}$ , que inclui um componente sistemático e um termo de erro:

$$U_{ij} = \mu_j + \varepsilon_{ij} \quad (1)$$

Somente o atributo do estado de saúde determina o componente sistemático da utilidade latente, por isso,  $\mu$  é indexado somente por  $j$ . Dessa forma, um dado estado de saúde tem um mesmo valor esperado de utilidade sombra entre todos os respondentes. Assume-se que as posições observadas para um conjunto de estados de saúde são correlacionadas a estes valores, que são distribuídos em torno de um nível médio para cada estado. Assim, um indivíduo somente dá uma alta posição para um estado de saúde com valor médio baixo, devido a sua variabilidade em relação à população, ou termo de erro (SALOMON, 2003; CRAIG *et al.*, 2009; McCabe *et al.*, 2006). Dada a tarefa de escolher o estado preferido em um grupo finito de diferentes estados, o indivíduo  $i$  ordenará o estado  $j$  acima que o estado  $k$  se  $U_{ij} > U_{ik}$  para todo  $j \neq k$ . Logo,

$$(\mu_j + \varepsilon_{ij}) > (\mu_k + \varepsilon_{ik}) \quad (2)$$

ou

$$(\mu_j - \mu_k) > (\varepsilon_{ik} - \varepsilon_{ij}) \quad (3)$$

Ao permitir elementos estocásticos no modelo, a probabilidade desta ordenação é dada por:

$$P(U_{ij} > U_{ik}) = P[(\varepsilon_{ik} - \varepsilon_{ij}) < (\mu_j - \mu_k)] \quad (4)$$

Assumindo que uma probabilidade nula é indistinguível de uma probabilidade extremamente pequena e que os termos de erro sejam independentes e identicamente distribuídos com uma distribuição valor extremo - EV-1 ou distribuição Weibull, tal que  $P(\varepsilon_{ij} \leq t) = \exp\{-\exp(-t)\}$ , então a razão de chances do estado  $j$  ser posicionado superiormente ao estado  $k$  é simplesmente igual a

$$\frac{P(\mu_j)}{P(\mu_k)} = \frac{\exp(\mu_j)}{\exp(\mu_k)} = \exp[(\mu_j - \mu_k)] \quad (5)$$

e a função de verossimilhança para o exercício completo de ordenação de um indivíduo é dada por:

$$L_i = \prod_{j=1}^J \left[ \frac{\exp\{\mu_j\}}{\sum_{k=1}^J \delta_{ijk} \exp\{\mu_k\}} \right] \quad (6)$$

onde  $j \neq k$  e  $\delta_{ijk}$  é um indicador da dominância do estado  $j$  sobre o estado  $k$  para o indivíduo  $i$ .

Como a razão de chances entre duas alternativas depende apenas de seus atributos, assume-se a **hipótese de independência das alternativas irrelevantes**, que implica que a adição de uma nova alternativa ou a alteração das características de uma terceira alternativa não afeta as probabilidades relativas entre as alternativas  $j$  e  $k$ . Esta hipótese dificilmente é mantida na presença de alternativas com alto grau de substitutabilidade. Um exemplo desta implicação é dado por McFadden (1974). Inicialmente, diante de duas alternativas de transporte, ônibus vermelho e carro, as pessoas escolhem cada uma delas com igual probabilidade, 0.5, de forma que a razão de chances da equação (5) seja unitária.

Supondo um terceiro modo de transporte, o ônibus azul, e assumindo que as preferências dos indivíduos por meio de locomoção não são influenciadas pela cor do ônibus, então as pessoas escolherão entre as três formas com igual probabilidade, 1/3. Assim, a fração de indivíduos que escolhem o carro cairia de 1/2 para 1/3, um resultado não muito realista. Na prática, pode-se agrupar ônibus vermelho e ônibus azul em uma mesma categoria, revelando que não existem outras diferenças além da cor – contudo, isto indicaria que para a manutenção da hipótese de independência das alternativas irrelevantes tornar-se-ia necessária a imposição de restrições no modelo logit condicional (WOOLDRIDGE, 2002).

Hausman e McFadden (1984) oferecem um teste baseado na verificação de que, se a hipótese se mantém, os parâmetros  $\beta$  podem ser consistentemente estimados através do modelo logit condicional utilizando apenas um subgrupo de alternativas. Assim, remove-se, aleatoriamente, a resposta referente a um estado de saúde dos dados de cada respondente. O modelo é reestimado usando a amostra reduzida, a fim de se comparar os coeficientes estimados com

aqueles encontrados a partir da amostra completa. O teste de *Hausman* é utilizado para operacionalizar esta análise.

No contexto dos exercícios realizados,  $\mu_j$  pode ser entendido como a valoração média de um estado de saúde e pode-se elaborar o modelo para expressar  $\mu$  como função das dimensões e níveis do sistema descritivo EQ-5D. O valor esperado da utilidade latente de cada estado de saúde é assumido ser uma função linear dos cinco domínios de saúde, sendo  $\theta$  um vetor de parâmetros desconhecidos e  $x_j$  o vetor das variáveis dummy para cada dimensão, que tomam valor igual a 1, se apresentam determinando nível de severidade e zero, caso contrário. Função de avaliação análoga ao modelo previamente estimado (ANDRADE *et al.*, 2013) a partir dos pesos obtidos pelo método da Troca de Tempo foi adotada para facilitar a comparação. Assim:

$$\mu_j = x_j\theta \quad (7)$$

Um conjunto de 10 variáveis *dummy* para cada dimensão e nível de severidade foi definido como se segue na TABELA 3. A variável dependente refere-se à posição dada por cada indivíduo aos estados de saúde no Exercício de Ordenação. O valor de  $\mu_j$  corresponde a valoração média dos estados de saúde, predita ou recomposta a partir dos coeficientes estimados  $\theta$  e das variáveis *dummy*  $x_j$ .

Uma matriz de variância robusta é necessária para se considerar a correlação entre as avaliações inseridas em cada questionário, devido aos efeitos não observados do respondente. Assim, a fim de se relaxar a hipótese de independência das observações, permite-se dependência dentro de cada cluster, através da correção da matriz de variância-covariância dos estimadores, pelos erros robustos de White e Ruber (WOOLDRIDGE, 2002). Além disso, a função de verossimilhança (6) configura-se na ausência de pareamento das ordenações dos estados de saúde. Para considerar a possibilidade de que um indivíduo classifique estados distintos em uma mesma posição, adotou-se a abordagem sugerida por Efron (1977), que substitui cada pareamento, por duas observações sintéticas, refletindo as duas ambíguas possibilidades de ordenação de um dado par de cartões. A cada um dos estados é atribuído um peso de 50% na função de verossimilhança (CRAIG *et al.*, 2009a).

O trabalho explora análises de equivalência entre os dados observados da EAV e do instrumento da Troca de Tempo, e os valores preditos do logit condicional gerados através das informações do exercício de ordenação. Os valores observados dos pesos populacionais, derivados a partir dos métodos cardinais, são calculados através de transformações <sup>9</sup> que restringem as utilidades, tal que os estados Morte e Saúde Perfeita apresentem valores iguais a 0 e 1, respectivamente.

**TABELA 3 – Descrição e Especificação das Variáveis do Modelo**

Variáveis	Descrição	Especificação
RANK	Posição do estado de saúde derivada a partir do Exercício de Ordenação	ordenação de 1 a 9, sendo atribuída posição 1 ao cartão associado ao melhor estado de saúde
TTO	Utilidade do estado de saúde derivada a partir do Exercício da Troca do Tempo	peso contínuo no intervalo [-1,1]
Mobilidade 2	Estado de saúde apresenta problema moderado na dimensão mobilidade	1 se mobilidade é nível 2; 0 caso contrário
Mobilidade 3	Estado de saúde apresenta problema grave na dimensão mobilidade	1 se mobilidade é nível 3; 0 caso contrário
Auto-cuidado 2	Estado de saúde apresenta problema moderado na dimensão auto-cuidado	1 se auto-cuidado é nível 2; 0 caso contrário
Auto-cuidado 3	Estado de saúde apresenta problema grave na dimensão auto-cuidado	1 se auto-cuidado é nível 3; 0 caso contrário
Atividades Habituais 2	Estado de saúde apresenta problema moderado na dimensão atividades habituais	1 se atividades habituais é nível 2; 0 caso contrário
Atividades Habituais 3	Estado de saúde apresenta problema grave na dimensão atividades habituais	1 se atividades habituais é nível 3; 0 caso contrário
Dor/Desconforto 2	Estado de saúde apresenta problema moderado na dimensão dor/desconforto	1 se dor/desconforto é nível 2; 0 caso contrário
Dor/Desconforto 3	Estado de saúde apresenta problema grave na dimensão dor/desconforto	1 se dor/desconforto é nível 3; 0 caso contrário
Ansiedade/Depressão 2	Estado de saúde apresenta problema moderado na dimensão ansiedade/depressão	1 se ansiedade/depressão é nível 2; 0 caso contrário
Ansiedade/Depressão 3	Estado de saúde apresenta problema grave na dimensão ansiedade/depressão	1 se ansiedade/depressão é nível 3; 0 caso contrário

Fonte: elaboração do autor.

Contudo, os resultados do modelo de efeitos aleatórios previamente estimados a partir do método TTO (ANDRADE *et al.*, 2013) também serão apresentados a fim de se compararem

<sup>9</sup> As fórmulas das transformações são apresentadas na subseção **Base de Dados**.

os coeficientes e as medidas de ajustes das métricas cardinais e ordinais<sup>10</sup>. Para se considerar a estrutura dos dados individuais em forma de painel, utilizou-se a modelagem de efeitos aleatórios (BANSBACK *et al.*, 2012b). O modelo padrão é definido como:

$$y_{ij} = (\beta x_{ij}) + \varepsilon_{ij} \quad (8)$$

Onde  $j = 1, 2, \dots, n$  representa o estado de saúde e  $i = 1, 2, \dots, m$  representa o indivíduo. A variável dependente  $y_{ij}$  corresponde ao valor da desutilidade ( $1 - TTO$ ) que o estado de saúde  $j$  gera para o indivíduo  $i$  e  $x_{ij}$  é um vetor de 10 variáveis explicativas binárias apresentadas na TABELA 3.

#### 4.2.2 O reescalonamento

O modelo logit condicional produz valores estimados definidos arbitrariamente em uma escala, de acordo com as hipóteses de identificação do modelo. Entretanto, uma medida genérica como o QALY, estimada através de métodos cardinais como Time Trade-Off e Standard Gamble, é ancorada pelos estados de saúde perfeita e morte, de modo a permitir comparações de intervenções de saúde distintas em uma escala comum. Assim, seguindo o processo descrito por Salomon (2003), a ordem em que os estados de saúde são classificados se mantém diante de uma transformação afim positiva das utilidades latentes, implicando na seguinte especificação geral da equação 1:

$$U_{ij} = \alpha(\mu_j + \varepsilon_{ij}) + \beta \quad (9)$$

Substituindo, na equação 9, a utilidade predita para um determinado estado de saúde (7) condicional aos valores dos parâmetros estimados no modelo, tem-se:

$$U_{ij} = \alpha x'_{ij} \hat{\theta} + \beta \quad (10)$$

No contexto das valorações dos estados de saúde, existem certas restrições conceituais dos possíveis valores dos parâmetros, dado que  $U_{ij} \in [-1, 1]$ . O parâmetro  $\beta$  representa o valor do

---

<sup>10</sup> Esta modelagem refere-se à principal contribuição de uma publicação anterior do grupo GEESC/UFMG, cujo objetivo era apresentar as preferências sociais por estados de saúde no Brasil (ANDRADE *et al.*, 2013).

estado caracterizado com os melhores níveis possíveis em todas as dimensões do EQ-5D (11111). Assim, quando um perfil não possui dificuldade em nenhuma das dimensões ( $x_j = 0$ ), o estado de saúde terá valor esperado de 1, e portanto, normaliza-se  $\beta = 1$ . Para os valores de  $\alpha$ , três alternativas são consideradas:

*Rank A - Normalização da escala para atribuir utilidade igual a zero para o pior estado de saúde (33333)*

O valor de  $\alpha$  pode ser definido de modo que, ao estado 33333, lhe seja atribuído utilidade igual a 0. Anteriormente ao processo de reescalonamento, o valor estimado para o estado 33333 é igual a soma de todos os coeficientes estimados no logit condicional referentes às variáveis *dummy* com dimensão de nível 3 - severo. Logo, substitui-se  $\beta = 1$  e  $\mu_j = \sum \hat{\theta}$  na equação 10,

$$\alpha \sum \hat{\theta} + 1 = 0 \quad (11)$$

$$\alpha = \frac{-1}{\sum \hat{\theta}} \quad (12)$$

*Rank B - Normalização da escala a fim de se estabelecer pareamento entre o valor do pior estado de saúde (33333) e sua utilidade estimada através do método da Troca de Tempo.*

Para transformar a escala tal que ao pior estado de saúde lhe seja atribuído o valor correspondente obtido a partir do método de Troca de Tempo, denotado por  $TTO_{33333}$ , tem-se que:

$$\alpha \sum \hat{\theta} + 1 = TTO_{33333} \quad (13)$$

$$\alpha = \frac{TTO_{33333} - 1}{\sum \hat{\theta}} \quad (14)$$

*Rank C - Normalização da escala para atribuir utilidade igual a zero para o estado Morte*

Nesta terceira opção de reescalonamento, as informações do Exercício de Ordenação referentes ao estado Morte são adicionadas à análise e um modelo estendido é estimado

incluindo todas as variáveis *dummy* além de uma variável categórica,  $\lambda$ , que toma valor igual a 1, se o estado é igual à Morte e 0, caso contrário. Logo,  $\lambda$  representa a utilidade estimada para o estado Morte a exemplo dos coeficientes usados para modelar as utilidades dos estados do sistema descritivo EQ-5D. Como definido anteriormente, o valor de  $\alpha$  é determinado por:

$$\alpha \hat{\lambda} + 1 = 0 \quad (15)$$

$$\alpha = -\frac{1}{\hat{\lambda}} \quad (16)$$

McCabe *et al.* (2006) e Brazier *et al.* (2009) consideram este procedimento de transformação o mais adequado, na medida em que mantém a âncora teórica do QALY, atribuindo valor nulo ao estado Morte, além de não impor o uso de dados externos ao próprio modelo, como o valor do pior estado de saúde hipotético gerado pelo método de Troca de Tempo.

#### 4.2.3 Medidas de Comparação

Após realizar as estimações, comparam-se as distribuições obtidas a partir dos diferentes métodos de obtenção de preferências. Em todos os estudos que incluem em seus protocolos tanto a EAV quanto a Ordenação, estes métodos apresentam correlação quase que perfeita, dado que a ordem de classificação dos estados de saúde permanece visível ao respondente quando da aplicação da escala. A versão brasileira do protocolo MVH realizada em Minas Gerais inova neste aspecto, já que, em todos os métodos de obtenção de pesos, os cartões impressos com os perfis de saúde são apresentados aos respondentes de forma aleatória, na expectativa de se obter uma evidência de robustez mais espontânea.

Os métodos TTO e EAV são designados a produzirem medidas cardinais de utilidade, mas podem ser transformados em valores ordinais no intuito de permitir comparações com as informações do Exercício de Ordenação. Para medir a correspondência entre estes posicionamentos, estima-se o *Rho de Spearman* para cada estado de saúde. Este coeficiente é uma medida de correlação não-paramétrica, isto é, avalia uma função monótona arbitrária que descreve a relação entre duas variáveis  $x$  e  $y$ , sem fazer nenhuma suposição sobre a distribuição de frequências destas (SIEGEL e CASTELLAN Jr., 2006). Seu cálculo é dado pela fórmula:

$$\rho = 1 - \frac{6\sum d_i^2}{n^3 - n} \quad (17)$$

onde  $d_i$  é a diferença entre cada posição ou posto do valor correspondente de  $x$  e  $y$  e  $n$  é o número de pares dos valores. Para amostras maiores que 20 observações, podemos utilizar a seguinte aproximação para a distribuição  $t$  de *Student*:

$$t = \frac{\rho}{\sqrt{\frac{(1-\rho^2)}{(n-2)}}} \quad (18)$$

Quanto às estimações do modelo de efeitos aleatórios e do modelo logit condicional, que são baseadas na mesma especificação, as comparações são realizadas utilizando-se o *Rho de Spearman* e o *Rho de Pearson*. Esta última medida de correlação, também chamada coeficiente de correlação produto-momento, requer a suposição de que a relação entre as variáveis seja linear e assume apenas valores entre -1 (correlação negativa perfeita) e 1 (correlação positiva perfeita). Seu cálculo é dado pela expressão:

$$\rho = \frac{COV(x, y)}{dp(x)dp(y)} \quad (19)$$

Quanto mais próximo de zero estiver o coeficiente de correlação, menos dependência linear existe entre as duas variáveis (SIEGEL e CASTELLAN Jr., 2006). Diferentemente do Erro Absoluto Médio, que representa a média dos erros cometidos por um modelo de previsão em relação aos valores observados e utilizados na própria estimação, neste trabalho utiliza-se a definição de Diferença Absoluta Média (DAM), que corresponde a diferença entre os valores preditos a partir dos diversos métodos e os valores observados gerados pela métrica da Troca de Tempo, que é o método comumente utilizado em avaliações. Assim, verifica-se o ajuste dos resultados oferecidos pelos métodos EAV e Ordenação, em relação àqueles encontrados no TTO, comparando-se duas unidades de medida distintas. Para fins de comparação com a modelagem padrão, realizada a partir dos dados do TTO (ANDRADE *et al.*, 2013), a DAM será comparada ao Erro Absoluto Médio encontrado para os valores preditos neste método.

Tanto a DAM quanto o número de estados de saúde com diferenças absolutas superiores a 0.05 e 0.10 são computados como medidas de qualidade de ajuste.

O cartão Morte é considerado um dos estados mais importantes nas análises de custo utilidade, na medida em que é utilizado como âncora na construção dos QALYs. Por isso, verifica-se a proporção de estados classificados como piores que a morte nos Exercícios de Ordenação, EAV e TTO, com o propósito de identificar se a posição da Morte varia entre os métodos de obtenção de preferências em saúde, com possíveis efeitos sobre os resultados. Também são realizadas comparações segundo o nível de severidade dos cartões. De acordo com Kind (2009), os estados são classificados com base na metodologia Estrutura Lógica Interna e Semiordenação Imposta, que se apóia na métrica da distância do bloco. Neste caso, contabiliza-se a distância de cada uma das dimensões do estado de saúde analisado em relação ao índice de saúde perfeita (11111). Assim, as distâncias para os estados 22222 e 33333 são 5 e 10, respectivamente. A classificação segue a seguinte definição:

*Brandos*: estados cujas distâncias estejam compreendidas entre 1 e 4; não possuam problemas de nível 3; e permitam, no máximo, 3 problemas de nível 2. A definição gera 25 estados classificados nesta categoria.

*Severos*: estados que não possuam problemas de nível 1 e que permitam, no mínimo, 2 problemas de nível 3, com exceção do pior estado de saúde hipotético (33333). Esta categoria inclui 25 estados de saúde.

*Moderados*: demais estados de saúde, com exceção do estado de saúde perfeita (11111), do pior estado de saúde hipotético (33333) e do estado Morte. Esta categoria compreende 50 estados de saúde.

Todas as análises estatísticas são computadas usando o programa Stata 12.0.

### **4.3 Inconsistências Lógicas**

Uma maneira de se avaliar a comparabilidade entre métodos de obtenção de preferências sociais é a capacidade de se produzir respostas consistentes com a racionalidade proposta pelo

estudo. Nesta seção, apresentam-se a definição e o cálculo de indicadores que refletem o nível de inconsistências das respostas, por método e por grupos de indivíduos.

#### 4.3.1 Definição de inconsistência

O sistema descritivo EQ-5D consiste em uma sequência de cinco dimensões, com três níveis de severidade cada uma. Essa sequência apresenta uma escala ordinal interna definida pelo próprio nível de severidade, de forma que, quanto maior a gravidade, pior é a condição de saúde da referida dimensão (DOLAN e KIND, 1996). Dessa forma, possuir problemas extremos de mobilidade é considerado pior que possuir apenas alguns problemas de mobilidade que, por sua vez, é pior que não possuir qualquer problema para caminhar. Como os estados de saúde são formados pela combinação de elementos para cada uma das cinco dimensões, estas descrições podem ser classificadas de acordo com a relação ordinal esperada entre elas. Contudo, não são todos os pares de estados de saúde passíveis de comparação.

Um par lógico ou comparável é aquele que permite o estabelecimento de uma conexão lógica de dominância entre os estados de saúde. A dominância entre os estados de saúde se verifica quando a mudança em cada dimensão entre os pares de estados respeitar sempre o mesmo sentido. No caso da mudança entre os pares de estados ocorrer somente em uma dimensão é suficiente analisar a diferença entre os níveis de severidade dessa dimensão. Por exemplo, o estado 12231 é logicamente dominado pelo estado 12221, já que para cada dimensão do primeiro, o nível de severidade é igual ou maior (pior) que o correspondente grau de severidade do segundo. Nesse caso ocorre mudança somente na dimensão de dor ou desconforto, passando da presença de problemas extremos para presença de alguns problemas.

Por outro lado, a relação entre os estados 12231 e um outro estado, 21231, não pode ser logicamente determinada, dado que os níveis das dimensões de 12231 não são superiores ou inferiores aos correspondentes níveis de 21231, de forma monotônica. Enquanto na dimensão de mobilidade, o primeiro estado de saúde é preferível ao segundo, no domínio referente aos cuidados pessoais, o sentido da relação de preferência é contrário. Assim, como os respondentes podem atribuir pesos distintos as diferentes dimensões, torna-se difícil estabelecer uma relação de dominância entre as duas condições de saúde.

A inconsistência ocorre quando um estado de saúde logicamente dominado, segundo a definição pré-estabelecida anteriormente, recebe uma valoração superior àquela atribuída a um estado de saúde dominante, em termos de qualidade de vida. Assim, se para o indivíduo  $i$  o estado 11212 é preferível ao estado 11112, este par de estados é considerado inconsistente, dado que o indivíduo parece não sustentar racionalidade em sua escolha. Para que o indivíduo respeite a hipótese pré-definida de ordinalidade, o valor que atribui ao estado  $j$  deve ser maior que o valor atribuído ao estado  $k$ , quando  $k$  é considerado logicamente pior, no mínimo, em uma das dimensões e não melhor, nos demais domínios.

A fim de definir os pares de estados logicamente comparáveis faz-se necessário formalizar a métrica proposta no instrumento EQ-5D, que se traduz em uma sequência de cinco algarismos. Cada algarismo representa o nível de severidade apresentado em cada uma das cinco dimensões de saúde. Faça o código de cada estado de saúde ser representado por  $L_{dMOB}L_{dCP}L_{dAH}L_{dDD}L_{dAD}$ , em que  $dMOB$ ,  $dCP$ ,  $dAH$ ,  $dDD$  e  $dAD$  representam as dimensões de mobilidade, cuidados pessoais, atividades habituais, dor ou desconforto e ansiedade ou depressão, respectivamente. Dessa forma,  $L_{dMOB}$ , um dos elementos ou algarismos da sequência, representa o nível de severidade referente à dimensão de mobilidade. No caso do estado de saúde 12213, os elementos tomam os seguintes valores:  $L_{dMOB}=1$ ,  $L_{dCP}=2$ ,  $L_{dAH}=2$ ,  $L_{dDD}=1$ ,  $L_{dAD}=3$ .

Sem perda de generalidade, faça  $h$  ser a dimensão do estado de saúde. Então, como existem três níveis de severidade – problemas brandos, problemas moderados e problemas extremos –  $L_{dh}$  assume valor entre 1 e 3. Assim, seja  $L_{dh}^j$  e  $L_{dh}^k$  os níveis de severidade da dimensão  $h$ , dos estados de saúde  $j$  e  $k$ , respectivamente. Então, o par de estados de saúde  $j$  e  $k$ ,  $C_{jk}$ , definido como

$$C_{jk} = \begin{cases} 1, & \text{se } L_{dh}^j - L_{dh}^k \geq 0 \quad \forall h \cup L_{dh}^j - L_{dh}^k \leq 0 \quad \forall h \\ 0, & \text{se } L_{dh}^j - L_{dh}^k \geq 0 \quad \cap \exists \text{ pelo menos um } h \mid L_{dh}^j - L_{dh}^k < 0 \end{cases} \quad (20)$$

é comparável ou lógico somente se assume valor unitário. Isso significa dizer que as diferenças entre cada um dos elementos dos dois estados de saúde, em suas respectivas dimensões, devem apresentar mesmo sinal – ou positivas, ou negativas. Quando se verificam, em um mesmo par, diferenças positivas e negativas, os estados de saúde não podem ser

considerados comparáveis ou lógicos. Segundo a hipótese pré-definida de ordinalidade, sabe-se que:

$$L_{dh}^j - L_{dh}^k \geq 0 \quad \forall h \rightarrow k \succ j \quad (21)$$

$$L_{dh}^j - L_{dh}^k \leq 0 \quad \forall h \rightarrow j \succ k \quad (22)$$

Isto é, se os níveis de severidade do estado  $j$  são superiores aos correspondentes níveis do estado  $k$ , para todas as dimensões de saúde  $h$ , então  $k$  domina  $j$ . Seja  $V_j^i$  o valor que o indivíduo  $i$  atribui ao estado de saúde  $j$ . Considerando monotonicidade estrita na escolha do agente, define-se inconsistência  $F$  do indivíduo  $i$  na avaliação do par de estados de saúde  $j$  e  $k$ , como:

$$F_{jk}^i = [(V_j^i > V_k^i) \cap (L_{dh}^j \geq L_{dh}^k)] \cup [(V_j^i < V_k^i) \cap (L_{dh}^j \leq L_{dh}^k)] \quad \forall h \quad (23)$$

Assim, a inconsistência na avaliação entre os pares de saúde  $j$  e  $k$  ocorre ou quando o valor que o indivíduo  $i$  atribui ao estado  $j$  é maior que àquele atribuído ao estado  $k$ , e simultaneamente, o estado de saúde  $j$  é dominado pelo estado  $k$ ; ou quando o valor atribuído ao estado  $j$  é menor que àquele atribuído ao estado  $k$ , e simultaneamente, o estado de saúde  $j$  domina o estado  $k$ . Definição alternativa também considera inconsistente o indivíduo que atribui um mesmo valor para estados de saúde comparáveis distintos. Assim, no caso da monotonicidade fraca ou não-estrita, tem-se:

$$F_{jk}^{*i} = [(V_j^i \geq V_k^i) \cap (L_{dh}^j \geq L_{dh}^k)] \cup [(V_j^i \leq V_k^i) \cap (L_{dh}^j \leq L_{dh}^k)] \quad \forall h \quad (24)$$

#### 4.3.2 Cálculo das inconsistências

A inconsistência interna será analisada utilizando-se dois indicadores: i) a porcentagem de respondentes que violam a ordem pré-estabelecida, no mínimo, em um par de estados de saúde; ii) a taxa de inconsistência individual, definida pela razão entre o número de pares

inconsistentes e o número de pares comparáveis, ou seja, que possibilitam uma relação de dominância lógica<sup>11</sup>, para cada indivíduo.

Para ilustrar o cálculo da taxa de inconsistência, imagine que, utilizando um dado método de obtenção de preferências, um indivíduo produz o seguinte ordenamento de estados de saúde: 31113, 33333, 32112 e 22221. Existem duas inconsistências nesta avaliação, dado que o pior estado de saúde hipotético 33333 nunca pode ser considerado melhor que os estados 32112 e 22221. Três pares de estados de saúde são comparáveis, pois tanto o estado 31113, quanto os estados 32112 e 22221 são preferíveis à condição 33333. Os pares 31113 e 32112, 31113 e 22221, e 32112 e 22221 não podem ser classificados como comparáveis segundo a definição apresentada. Dessa forma, o respondente realizou duas de três inconsistências possíveis, fornecendo uma taxa de inconsistência igual a 2/3. Esta taxa varia entre 0 (ausência de pares inconsistentes) e 1 (todos os pares lógicos são inconsistentes).

Através do conjunto de 243 estados de saúde possíveis gerados pelo sistema descritivo EQ-5D, pode-se formar  $C_{243,2} = 29.403$  pares, dos quais 7.533 (25,6%) são comparáveis por construção, ou seja, possuem relação de dominância lógica entre seus domínios (CRAIG e RAMACHANDRAN, 2006). Algumas inovações sugeridas pelo Grupo EuroQol (KIND, 2009) foram adotadas na pesquisa. Uma delas constitui-se na obtenção de 102 estados de saúde, de forma direta, e a utilização de 26 blocos fixos, cada um composto por seis estados de saúde pré-definidos, de forma a balancear os três níveis de severidade possíveis. Assim, um estado de saúde somente é avaliado com aqueles estados que pertencem aos grupos em que se encontra. Este desenho impõe uma quantidade ainda menor de pares comparáveis, já que o número de estados de saúde expostos simultaneamente a uma mesma avaliação, é reduzido.

---

<sup>11</sup> Um extrato da matriz  $M_{op}$  de dominância lógica entre todos os 102 estados de saúde analisados encontra-se no Apêndice B. O elemento  $x_{op}$ , da linha  $o$  e coluna  $p$  da matriz  $M_{op}$ , se refere à ordem lógica pré-estabelecida entre um par de estados de saúde qualquer  $j$  e  $k$ , e toma valores:

$$x_{op} = \begin{cases} 1, & \text{se } o \succ p \\ -1, & \text{se } o \prec p \\ 0, & \text{se } C_{jk} = 0 \end{cases}$$

Além disso, o cartão Morte não foi considerado no cálculo das inconsistências. Primeiro, porque a Morte é uma das âncoras utilizadas no exercício de Troca do Tempo e, portanto, não apresenta avaliações para este estado. Segundo, e mais importante, permitir que um estado de saúde classificado como pior que a morte seja considerado como inconsistência viola o pressuposto básico ou o objetivo inicial do método baseado na troca - *choice-based*. Neste sentido, dado que cada indivíduo avaliou um total de oito estados de saúde e o cartão Morte, verifica-se que o número de pares de estados comparáveis para cada respondente não ultrapassou o máximo de  $C_{8,2} = 28$ , considerando que, dificilmente, todos os oito estados de saúde possuam relação de dominância lógica.

Diante da importância de se considerar a qualidade e a representatividade dos dados em análise, verifica-se a possibilidade de diferenças na produção de inconsistências pelos três diferentes métodos de avaliação. Essas diferenças serão verificadas a partir de estatísticas de tendência central como média e mediana da taxa de inconsistência, além da proporção de indivíduos inconsistentes. Além disso, busca-se analisar se características individuais como educação, condição socioeconômica do domicílio e idade do respondente possuem algum efeito sobre a geração de escolhas inconsistentes. As diferenças por subgrupos de respondentes serão avaliadas, por meio da média das taxas de inconsistência e do modelo logit para respostas binárias apresentado na próxima subseção.

Diferenças de médias são obtidas a partir dos testes t de Student e os métodos não-paramétricos de Mann-Whitney e Kruskal-Wallis. Todos os testes requerem que a variável em análise seja medida em escala ordinal ou numérica e testam a hipótese nula de que as k amostras provêm da mesma população. Os testes t de Student e Mann-Whitney são usados para testar se duas amostras independentes foram extraídas de populações com médias iguais. Contudo, o segundo serve como alternativa ao primeiro, principalmente, quando a variável populacional analisada não segue uma distribuição normal e/ou as amostras forem pequenas. Quanto ao teste Kruskal-Wallis, este é extremamente útil para decidir se k amostras ( $k > 2$ ) independentes provêm de populações com médias iguais (SIEGEL e CASTELLAN Jr., 2006).

Por fim, os respondentes serão agrupados de acordo com o número de inconsistências gerados em suas avaliações: 0, 1, 2, 3 e mais que 3 inconsistências. Considerando a possibilidade de um erro aleatório na avaliação individual, acredita-se ser razoável que um indivíduo cometa apenas uma inconsistência. Neste sentido, inclui-se o grupo de respondentes com 0 ou 1

violação da ordem lógica. Para cada um desses grupos, são calculadas as tarifas médias de todos os estados de saúde. No caso do TTO e da EAV, obtém-se os valores observados médios. Quanto ao Exercício de Ordenação, a estimação de pesos cardinais a partir de informações ordinais se dá a partir do modelo logit condicional apresentado anteriormente, para cada uma das seis subamostras. Para captar as alterações de pesos estimados impostas pelas inconsistências, os coeficientes encontrados pelo modelo logit são comparados àqueles estimados para o método da Troca de Tempo através do modelo de efeitos aleatórios.

#### 4.3.3 Modelo Logit

Seja a variável  $y_i$ , com  $y = 0$  ou  $y = 1$ , uma manifestação observável da variável latente  $y_i^*$ , determinada por um conjunto de características  $x_i$

$$y_i^* = x_i\beta + u_i \quad (25)$$

e a regra de determinação de  $y_i$  em função de  $y_i^*$  definida abaixo

$$y_i = \begin{cases} 1, & \text{se } y_i^* > 0 \\ 0, & \text{se } y_i^* \leq 0 \end{cases} \quad (26)$$

Então, quanto maior for o valor da variável latente, maior será a probabilidade de escolha da alternativa  $y_i = 1$ :

$$P_i = \text{Prob}(y_i = 1) = \text{Prob}(y_i^* > 0) = \text{Prob}(x_i\beta + u_i > 0) = \text{Prob}(u_i > -x_i\beta) \quad (27)$$

e, por conseguinte

$$P_i = \text{Prob}(y_i = 0) = \text{Prob}(u_i \leq -x_i\beta) \quad (28)$$

Sendo  $F(\cdot)$  a distribuição acumulada de  $u$ , e assumindo que  $u$  seja independente de  $x$  e tenha distribuição logística padrão, tem-se:

$$\text{Prob}(y_i = 0) = F(-x_i\beta) \quad (29)$$

$$\text{Prob}(y_i = 1) = 1 - F(-x_i\beta) = F(x_i\beta) \quad (30)$$

A meta principal é estimar o efeito parcial de  $x_j$  sobre a probabilidade de sucesso  $Prob(y_i = 1)$ . Contudo, os coeficientes integrados no vetor  $\beta$  não têm a interpretação usual dos modelos de regressão linear, justamente devido à natureza não-linear de  $F(\cdot)$ . Assim,

$$\frac{\Delta p(x)}{\Delta x_j} = \frac{\Delta F(x_i\beta)}{\Delta(x_i\beta)} \frac{\Delta(x_i\beta)}{\Delta x_j} = f(x_i\beta) \beta \quad (31)$$

em que  $f(\cdot)$  designa a função densidade probabilidade correspondente à função de distribuição acumulada  $F(\cdot)$ , que é estritamente crescente. Dessa forma,  $f(\cdot) > 0$ , de modo que o efeito parcial de  $x_j$  sobre  $p(x)$  dependa de  $x$  devido ao fator  $f(x_i\beta)$  e sempre tenha o mesmo sinal de  $\beta$ . Contudo, os efeitos relativos de duas variáveis explicativas contínuas não dependem de  $x_i$ : a razão dos efeitos parciais de  $x_j$  e  $x_k$  é  $\beta_j/\beta_k$  (WOOLDRIDGE, 2002). Como  $u_i$  segue uma distribuição logística:

$$P(y_i = 1) = \frac{1}{1 + \exp^{-x\beta}} \quad (32)$$

$$p(y_i = 0) = 1 - \frac{1}{1 + \exp^{-x\beta}} = \frac{\exp^{-x\beta}}{1 + \exp^{-x\beta}} \quad (33)$$

$$\frac{P(y_i = 1)}{P(y_i = 0)} = \exp^{x\beta} \quad (34)$$

$$\ln\left(\frac{P(y_i = 1)}{P(y_i = 0)}\right) = x\beta \quad (35)$$

O logaritmo natural do quociente entre a probabilidade de sucesso e a probabilidade de fracasso, também conhecido como a razão de chances, revela quantas vezes é provável que ocorra o dado fenômeno frente a que não ocorra. A razão de chances sempre será maior que zero e se o valor obtido for superior à unidade, a probabilidade de sucesso é maior que a de fracasso.

Caso  $x_j$  seja uma variável explicativa binária, o efeito parcial de alterar  $x_j$  de zero para um, mantendo-se fixas as demais variáveis, será:

$$F(\beta_0 + \beta_1 x_1 + \dots + \beta_j + \dots + \beta_k x_k) - F(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (36)$$

Os parâmetros do vetor  $\beta$  são estimados através do método de Máxima Verossimilhança, de forma a considerar a heterocedasticidade em  $Var(y/x)$ . Dado uma amostra aleatória de tamanho  $n$  e uma variável resposta com distribuição de probabilidade do tipo Bernoulli, a função de máxima verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^n \left\{ [\text{Pr ob}(y_i=0)]^{1-y_i} [\text{Pr ob}(y_i=1)]^{y_i} \right\} \quad (37)$$

Ou, substituindo por (29) e (30),

$$L(\beta) = \prod_{i=1}^n \left\{ [1 - F(x_i\beta)]^{1-y_i} [F(x_i\beta)]^{y_i} \right\} \quad (38)$$

Percebe-se que quando  $y = 1$ , obtém-se  $F(x_i\beta)$  e quando  $y = 0$ , obtém-se  $1 - F(x_i\beta)$ . Os estimadores de máxima verossimilhança do vetor  $\beta$  são funções de valores amostrais que maximizam  $L(\beta)$  em relação a  $\beta$  (CAMERON e TRIVEDI, 2005). Contudo, maximizar a função  $L(\beta)$  torna-se uma tarefa mais simples ao tomar a função logarítmica de verossimilhança:

$$\ell(\beta) = \sum_{i=1}^n \left\{ (1 - y_i) \ln[1 - F(x_i\beta)] + y_i \ln[F(x_i\beta)] \right\} \quad (39)$$

Diferenciando com respeito a  $\beta$ , tem-se a condição de primeira ordem abaixo:

$$\sum_{i=1}^n \left\{ \frac{y_i}{F_i} F_i' x_i - \frac{1 - y_i}{1 - F_i} F_i' x_i \right\} = 0 \quad (40)$$

Dada a não-linearidade do sistema composto pelas c.p.o., não existe solução explícita para  $\hat{\beta}$ . No entanto, a partir do procedimento iterativo Newton-Raphson, realiza-se a otimização numérica que permite a identificação do máximo global (CAMERON e TRIVEDI, 2005).

Na regressão logística, a variável resposta dicotômica  $y_i$  assume valor igual à unidade se o indivíduo  $i$  avalia, no mínimo, um par de estados de saúde de forma a violar a ordenação lógica pré-estabelecida; e zero, caso o respondente seja totalmente consistente em suas decisões. Portanto, as probabilidades de sucesso estão associadas a agentes inconsistentes. As variáveis independentes, listadas na TABELA 4, referem-se às características individuais dos respondentes, além de variáveis relativas ao procedimento de pesquisa.

**TABELA 4 – Descrição e Especificação das Variáveis do modelo**

Variáveis	Descrição	Especificação
Inconsistente	Variável dependente relacionada à inconsistência do respondente em suas avaliações	1 se o indivíduo gera pelo menos uma inconsistência 0 se o indivíduo é consistente
Educação	Nível de escolaridade do respondente: Primário Incompleto Primário Completo Fundamental Completo Médio Completo Superior Completo	1 se Primário Incompleto (referência) 2 se Primário Completo 3 se Fundamental Completo 4 se Médio Completo 5 se Superior Completo
Idade	Idade do respondente em anos	contínua
Classe Socioeconômica	Classe socioeconômica do domicílio do respondente: Classe A Classe B Classe C Classe D-E	Pontuação no Critério Brasil: 35 a 46 pontos - Classes A1 e A2 (referência) 23 a 34 pontos - Classes B1 e B2 14 a 22 pontos - Classes C1 e C2 0 a 13 pontos - Classes D e E
Homem	Sexo do respondente	1 se homem; 0 se mulher (referência)
Estado Civil	Estado civil do respondente Casado Viúvo Divorciado/separado Solteiro	1 se casado (referência) 2 se viúvo 3 se divorciado/separado 4 se solteiro
Trabalhou na área da saúde	Experiência de trabalho do respondente na área da saúde	1 se trabalhou; 0 se não trabalhou (referência)
Possui plano de saúde	Posse de plano de saúde pelo respondente	1 se possui; 0 se não possui (referência)
Saúde auto-avaliada	Como o respondente avalia sua própria saúde: Muito bom ou bom Regular Ruim ou muito ruim	1 se muito bom ou bom (referência) 2 se regular 3 se ruim ou muito ruim
Feliz	Como o respondente avalia seu nível de felicidade: Feliz ou muito feliz Infeliz ou não muito feliz	1 se feliz ou muito feliz 0 se infeliz ou não muito feliz (referência)
Tempo de Entrevista	Tempo de duração da entrevista	contínua
Dificuldade no TTO	Como o respondente avalia o nível de dificuldade do método da Troca de Tempo	1 se difícil ou muito difícil 0 se fácil ou muito fácil (referência)

Fonte: elaboração do autor.

A variável socioeconômica utilizada foi o “Critério Brasil”, um instrumento de segmentação econômica que diferencia a população de acordo com dois fatores: o grau de instrução do chefe de família e as características domiciliares, que são definidas pela posse de itens, tais como, geladeira, televisão em cores, automóvel, empregada mensalista, etc. O critério atribui pontos em função de cada uma destas características e a soma destes pontos – que varia entre 0 e 46 – define em qual “classe econômica” o domicílio se enquadra (ABEP, 2013). Há uma correspondência entre as faixas de pontuação do critério e estas classes, que são definidas por A1, A2, B1, B2, C1, C2, D e E, sendo a classe A1 aquela que abrange os domicílios com pontuação entre 42 e 46 (mais ricos) e a classe E a que abarca os domicílios entre 0 e 7 pontos (mais pobres). Quatro grupos sínteses são construídos a partir das classes socioeconômicas, a saber, os grupos A, B, C, D-E.

O cálculo do número de inconsistências e a construção das taxas de inconsistência por indivíduo foram realizados através do software livre Dev-C++. Todas as análises estatísticas são computadas usando o programa Stata 12.0.

## 5. RESULTADOS

Esta seção é composta por duas subseções. Na primeira, são apresentados os dados cardinais estimados a partir de informações ordinais oferecidas pelo Exercício de Ordenação e a comparação entre os pesos médios estimados através dos três métodos de obtenção de preferências analisados no estudo: Ordenação, Escala Analógica Visual e o exercício da Troca de Tempo. Na segunda subseção, analisam-se as inconsistências lógicas, tanto por métodos de avaliação, quanto por características individuais dos respondentes, além de se verificar o efeito das inconsistências sobre os valores médios estimados.

### 5.1 Resultados do Modelo Logit Condicional

Um achado que se mostra interessante é a alta, mas não perfeita correlação entre os valores resultantes do exercício de ordenação e o *ranking* baseado na escala analógica visual. Diferentemente dos resultados apresentados por Craig *et al.* (2009), onde as duas variáveis são praticamente colineares para maioria dos países analisados, o coeficiente de *Spearman* registra valor igual a 0.696. Dessa forma, parece importante apresentar os estados de saúde de forma aleatória ao entrevistado a fim de que a verificação de correlação seja espontânea. Como esperado, os resultados da ordenação tradicional e daquela derivada através do TTO apresentaram menor correlação, com coeficiente de *Spearman* igual a 0.564.

Uma distinção encontrada entre os métodos, refere-se à posição ocupada pelo cartão Morte nos exercícios. Enquanto o número de estados classificados como piores que a morte por respondente nos instrumentos de ordenação e EAV foi de 0.24 e 0.25, respectivamente, este número chega a 1.91 para o método TTO. Dessa forma, enquanto na abordagem ordinal e na escala, a cada quatro respondentes, apenas um classifica um único estado de saúde como pior que a morte, na técnica baseada na troca, cada respondente classifica, em média, quase dois estados como piores que a morte. Como as transformações arbitrárias que limitam os valores dos estados piores que a morte podem afetar as diferenças cardinais, mas não mitigar as distâncias ordinais entre as métricas, verifica-se uma diferença consistente na posição do estado Morte, sendo este posicionado inferiormente no Exercício de Ordenação. Cerca de 12% e 11% dos respondentes classificam algum estado de saúde como pior que a morte no exercício de ordenação e na escala analógica visual, respectivamente. Quanto ao método da Troca de Tempo, este percentual chega a aproximadamente 74%.

Os resultados encontrados para o caso brasileiro confirmaram os achados internacionais no que se refere às estimações de dados cardinais a partir de informações ordinais. Todos os valores preditos são altamente correlacionados, com grande parte dos coeficientes de correlação acima de 0.95. Os maiores acordos se dão entre os pesos observados e preditos do TTO e entre os valores esperados do modelo logit condicional segundo diferentes tipos de reescalonamento.

**TABELA 5 – Correlações de *Pearson* e *Spearman* dos valores esperados por método**

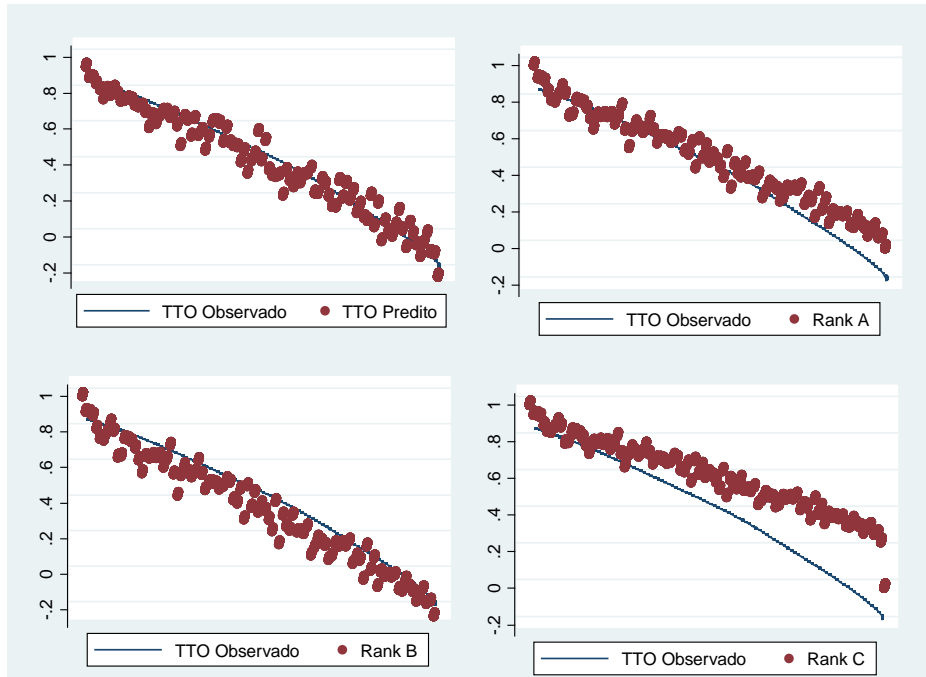
Correlação de Pearson						
	TTO Observado	TTO Predito	EAV	Rank A	Rank B	Rank C
TTO Observado	1					
TTO Predito	0,9866	1				
EAV	0,9353	0,9484	1			
Rank A	0,9741	0,9845	0,9566	1		
Rank B	0,9741	0,9845	0,9566	1	1	
Rank C	0,9729	0,9832	0,9569	1	1	1
Correlação de Spearman						
	TTO Observado	TTO Predito	EAV	Rank A	Rank B	Rank C
TTO Observado	1					
TTO Predito	0,9867	1				
EAV	0,945	0,9547	1			
Rank A	0,973	0,9826	0,9596	1		
Rank B	0,973	0,9826	0,9596	1	1	
Rank C	0,9716	0,9814	0,9606	0,9998	0,9998	1

Fonte: elaboração do autor.

Nota: Todos os valores são significativos a 1%.

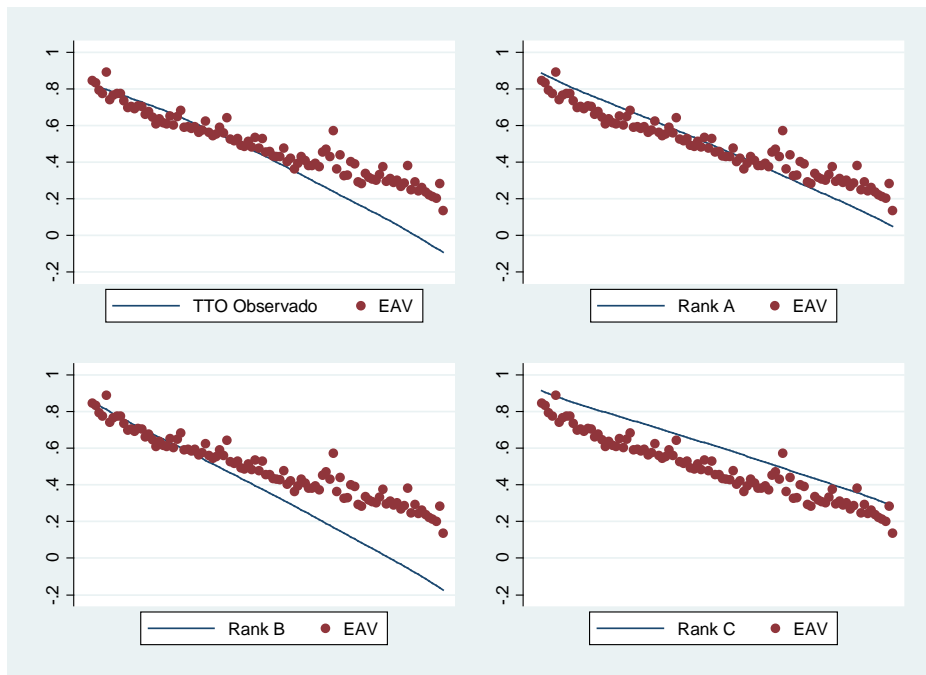
As três alternativas calculadas para o modelo logit condicional foram fortemente correlacionadas com os valores médios observados do TTO: o coeficiente de *Pearson* foi igual a 0.9741 para as opções A e B – que por definição, são transformações lineares uma da outra, e 0.9729 para a opção C – que desvia desta relação linear na medida em que um modelo estendido foi estimado com a inclusão da variável morte. Os resultados da Escala Analógica Visual e do TTO apresentaram a menor correlação linear, 0.935. A FIGURA 3 ilustra o ajuste entre os valores observados do TTO e os valores esperados do modelo logit a partir das diferentes formas de reescalonamento. A distinção entre os valores preditos do Rank B e do Rank C destaca a diferença no número de estados de saúde classificados como piores que a morte no exercício de ordenação vis-à-vis às respostas no TTO, refletindo a evidência de que os indivíduos possuem maior dificuldade em classificar estados de saúde como piores que a morte na métrica ordinal.

**FIGURA 3 – Valores esperados do método da Troca de Tempo e do Exercício de Ordenação segundo diferentes formas de reescalonamento**



Fonte: elaboração do autor.

**FIGURA 4 – Valores observados da EAV e do TTO e valores preditos no modelo logit condicional**



Fonte: elaboração do autor.

Além disso, a estimativa do coeficiente para o estado Morte, necessária para normalizar os pesos na escala 0-1, é afetada pela variabilidade das frequências com que o estado Morte é avaliado dominante em relação aos demais. Como poucas vezes se realiza troca entre tempo e qualidade de vida no Exercício de Ordenação, as escolhas envolvendo a morte são determinísticas, contrariando o pressuposto de aleatoriedade do modelo de utilidade aleatória. A FIGURA 4 mostra como a EAV sobre-estima os valores preditos para os estados severos, quando comparada ao TTO, ao Rank A e ao Rank B; e subestima os pesos para todos os estados em relação ao Rank C. A TABELA 6 apresenta os resultados do modelo logit condicional, segundo o reescalonamento de melhor ajuste (Rank B), e os resultados do modelo de efeitos aleatórios, para o método da Troca de Tempo<sup>12</sup>.

**TABELA 6 – Resultados do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo)**

Variáveis	Coefficientes TTO	dp	Coefficientes Rank	dp
Mobilidade 2	0,127***	0,007	1,142***	0,029
Cuidados Pessoais 2	0,121***	0,007	0,688***	0,026
Atividades Habituais 2	0,095***	0,007	0,503***	0,026
Dor/Desconforto 2	0,067***	0,007	0,436***	0,023
Ansiedade/Depressão 2	0,061***	0,007	0,404***	0,025
Mobilidade 3	0,403***	0,008	2,307***	0,046
Cuidados Pessoais 3	0,246***	0,008	1,052***	0,031
Atividades Habituais 3	0,205***	0,008	0,869***	0,031
Dor/Desconforto 3	0,200***	0,007	0,751***	0,027
Ansiedade/Depressão 3	0,113***	0,007	0,661***	0,027
n. estados		102		102
Diferença Absoluta Média		0,035		0,073
No. >0.05		25		64
No. >0.10		2		29
rho de Pearson		0,987		0,974
rho de Spearman		0,987		0,973

Fonte: elaboração do autor.

\*\*\* Significante a 1%.

Nota: Os coeficientes estimados pelo modelo logit condicional foram utilizados no Rank A e no Rank B. As estatísticas de ajuste referem-se ao modelo de normalização Rank B.

Todos os coeficientes são positivos e significativos a 1%. Assim como os coeficientes do método TTO, o modelo logit condicional apresenta comportamento esperado quanto aos sinais dos parâmetros estimados. No modelo de efeitos aleatórios, verificou-se um aumento

<sup>12</sup> Apesar do pior ajuste, os coeficientes do Rank C são bastante similares aos encontrados pelo Rank B. O Apêndice C apresenta os coeficientes e as estatísticas de ajuste para o modelo de normalização baseado na morte.

monotônico no valor da desutilidade à medida que se elevou o nível de severidade em todas as dimensões. Para os dados ordenados, o melhor estado de saúde hipotético é classificado na primeira posição. Dessa forma, quanto pior o estado de saúde, maior sua classificação na ordenação. De fato, verificou-se um aumento dos coeficientes estimados correspondentes à presença de algum problema nas dimensões de saúde. O maior decréscimo de utilidade foi observado para problemas de mobilidade severo, enquanto a menor desutilidade foi observada para problemas moderados no domínio ansiedade/depressão. Apesar destes decréscimos serem sempre maiores no modelo de Ordenação, quando comparados ao método cardinal, verifica-se equivalência entre os modelos no que se refere à importância relativa das dimensões e níveis de severidade. Neste sentido, o exercício ordinal parece representar uma alternativa válida para se mensurar preferências, dado que apresenta resultados com comportamentos similares aos encontrados pela abordagem cardinal.

O teste da hipótese de independência das alternativas irrelevantes é reportado na TABELA 7. Os resultados fornecem a evidência de que a hipótese do modelo logit se mantém. A partir da análise do *p*-valor do teste de *Hausman*, não se pode rejeitar a hipótese nula de igualdade dos coeficientes.

**TABELA 7 - Teste de *Hausman* para a Hipótese de Independência das Alternativas Irrelevantes**

Variáveis	Amostra Completa	dp	Amostra Reduzida	dp
Mobilidade 2	1,143	0,029	1,148	0,031
Cuidados Pessoais 2	0,689	0,026	0,698	0,028
Atividades Habituais 2	0,503	0,026	0,510	0,030
Dor/Desconforto 2	0,436	0,023	0,431	0,026
Ansiedade/Depressão 2	0,405	0,025	0,412	0,028
Mobilidade 3	2,307	0,046	2,326	0,038
Cuidados Pessoais 3	1,052	0,031	1,043	0,031
Atividades Habituais 3	0,870	0,031	0,877	0,032
Dor/Desconforto 3	0,752	0,027	0,757	0,029
Ansiedade/Depressão 3	0,662	0,027	0,661	0,029
Teste Hausman		X <sup>2</sup> = 6,48		p-value = 0,7731

Fonte: elaboração do autor.

Nota: Todos os coeficientes foram significativos a 1%.

O método da Troca de Tempo apresenta uma performance superior à métrica derivada de dados ordinais, em todas as medidas de qualidade de ajuste. Entretanto, este resultado não surpreende, dado que todas as comparações são realizadas com base nos valores observados do TTO, que é o método considerado referência pela sua grande utilização em análises de custo-utilidade. Assim, os resultados preditos com base nos valores observados do TTO

apresentaram ajuste superior aos encontrados pelo modelo logit condicional que, apesar de avaliar os mesmos indivíduos, baseia-se em um diferente conjunto de dados. Neste sentido, averiguou-se a similaridade entre as estimativas de dois métodos totalmente distintos. O erro médio absoluto próximo a 0.035, encontrado pelo modelo de efeitos aleatórios, é menor que o limite comumente pré-especificado de 0.05. Contudo, dada a diferença da natureza das informações sob as quais os modelos estão baseados, o modelo logit condicional apresenta uma performance de ajuste aceitável, com diferença média absoluta inferior a 10% para o Rank A e o Rank B. A FIGURA 3 confirma estas similaridades.

As porcentagens de estados de saúde com diferenças médias absolutas superiores a 0.05 e 0.10 são altas quando comparadas ao método cardinal. Na melhor das transformações, 64% dos cartões apresentaram desvio maior que 5%. O reescalonamento que define morte com utilidade igual a zero, apresentou o pior ajuste, seguido pelos resultados da EAV. Como descrito na TABELA 8, com exceção do método TTO, que apresentou pior ajuste nos cartões moderados, todos os modelos obtiveram maiores diferenças médias absolutas para estados de saúde severos. Este fato, supostamente está relacionado à maior facilidade com que o método TTO possibilita a troca entre tempo e qualidade de vida, registrando menores valores para estados severos, quando comparado às demais métricas de avaliação.

**TABELA 8 – Diferença Absoluta Média (DAM) total e por nível de severidade do estado de saúde**

	Total				
	TTO	EAV	Rank A	Rank B	Rank C
DAM	0,035	0,118	0,079	0,073	0,189
No. >0.05 (%)	25	70	66	64	95
No. >0.10 (%)	2	47	33	29	73
<b>Brandos (25)</b>					
DAM	0,021	0,050	0,039	0,063	0,076
No. >0.05 (%)	1	10	8	15	22
No. >0.10 (%)	0	4	0	6	7
<b>Moderados (50)</b>					
DAM	0,043	0,098	0,084	0,074	0,196
No. >0.05 (%)	18	36	35	31	48
No. >0.10 (%)	2	19	18	14	41
<b>Severos (25)</b>					
DAM	0,035	0,219	0,105	0,085	0,301
No. >0.05 (%)	6	23	22	18	24
No. >0.10 (%)	0	23	14	9	24

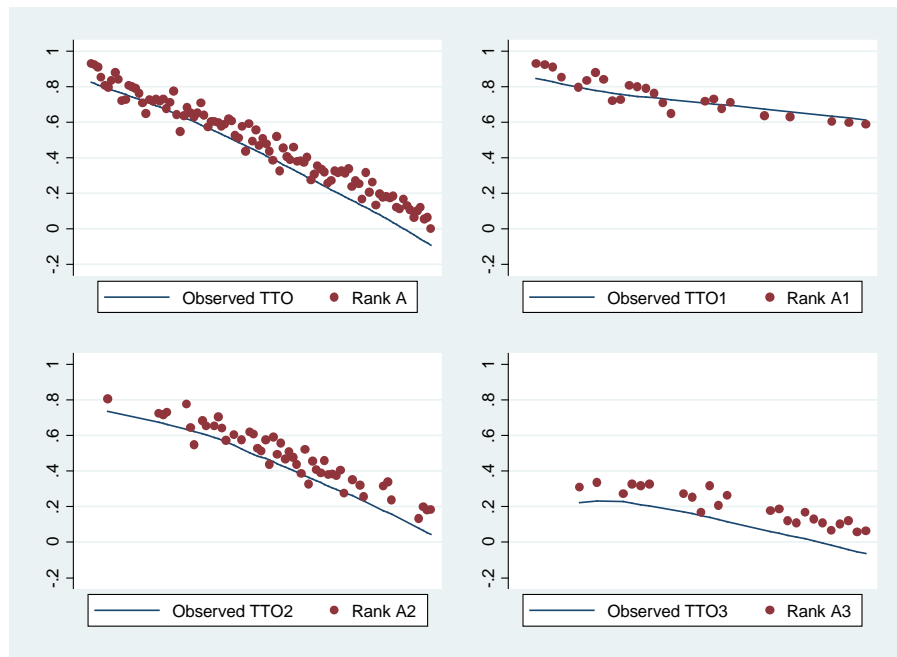
Fonte: elaboração do autor.

Dos 25 cartões classificados como severos, quase a totalidade (96%) superou o limite de 10% em termos de diferença média absoluta para o Rank C. No que se refere ao modelo de melhor

ajuste - Rank B, dos sete estados com maior discrepância entre o valor predito na ordenação e o valor observado no TTO, seis estados possuíam dois problemas severos em alguma dimensão, três estados possuíam três problemas moderados, e todos possuíam pelo menos um atributo com nível de severidade extremo. As FIGURAS 5, 6, 7 e 8 ilustram como os piores ajustes se dão para os estados de saúde mais precários. Enquanto os pesos esperados com base nas informações da EAV, do Rank A e do Rank C implicam em maiores valorações à medida que aumenta os problemas dos perfis de saúde, o modelo logit condicional subestima os valores para estados classificados como severos no Rank B.

As relações entre TTO e a EAV e entre TTO e os três modelos de *ranking* apresentam uma não-linearidade, tal que a inclinação é mais suave para estados de saúde severos, e mais íngreme, para os perfis brandos (FIGURAS 9 e 10). Esta possível curvatura na relação entre os valores preditos sugere uma forma funcional quadrática para a função valor, geralmente estimada a fim de se transformar pesos obtidos através do método EAV em utilidades TTO. A relação entre os valores preditos a partir dos dados ordinais e aqueles observados na EAV parece ser linear e confirma os achados para os países europeus.

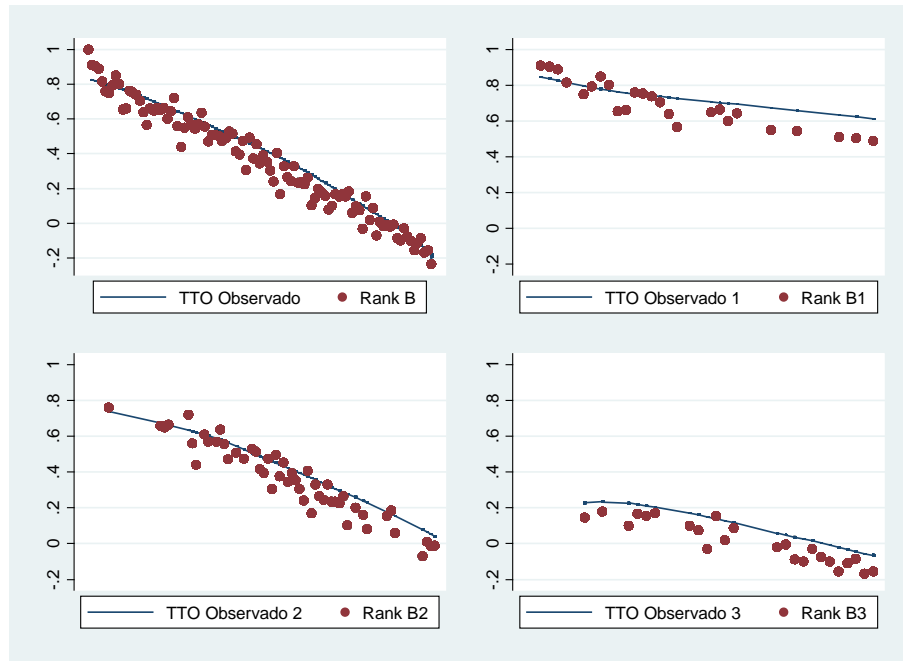
**FIGURA 5– Valores preditos do Rank A e observados do TTO por nível de severidade do estado de saúde**



Fonte: elaboração do autor.

Nota: (1) estados brandos (2) estados moderados (3) estados severos.

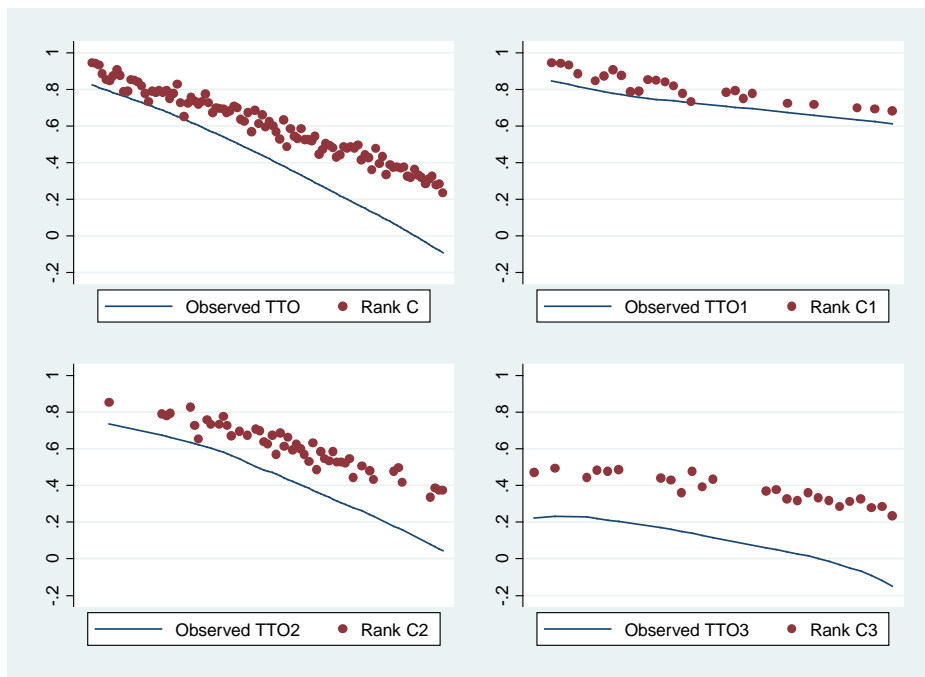
**FIGURA 6– Valores preditos do Rank B e observados do TTO por nível de severidade do estado de saúde**



Fonte: elaboração do autor.

Nota: (1) estados brandos (2) estados moderados (3) estados severos.

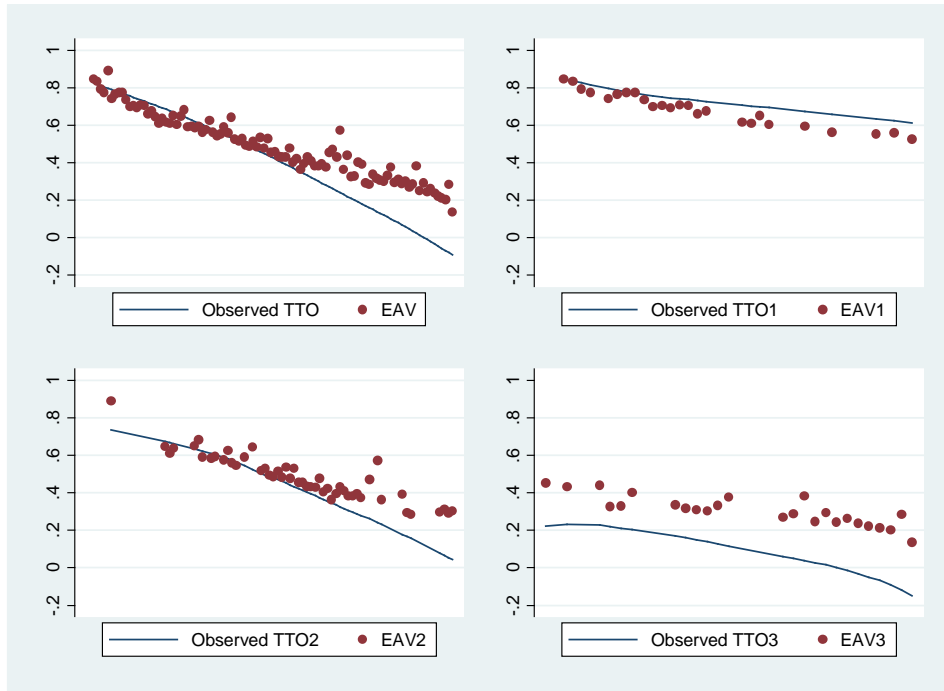
**FIGURA 7– Valores preditos do Rank C e observados do TTO por nível de severidade do estado de saúde**



Fonte: elaboração do autor.

Nota: (1) estados brandos (2) estados moderados (3) estados severos.

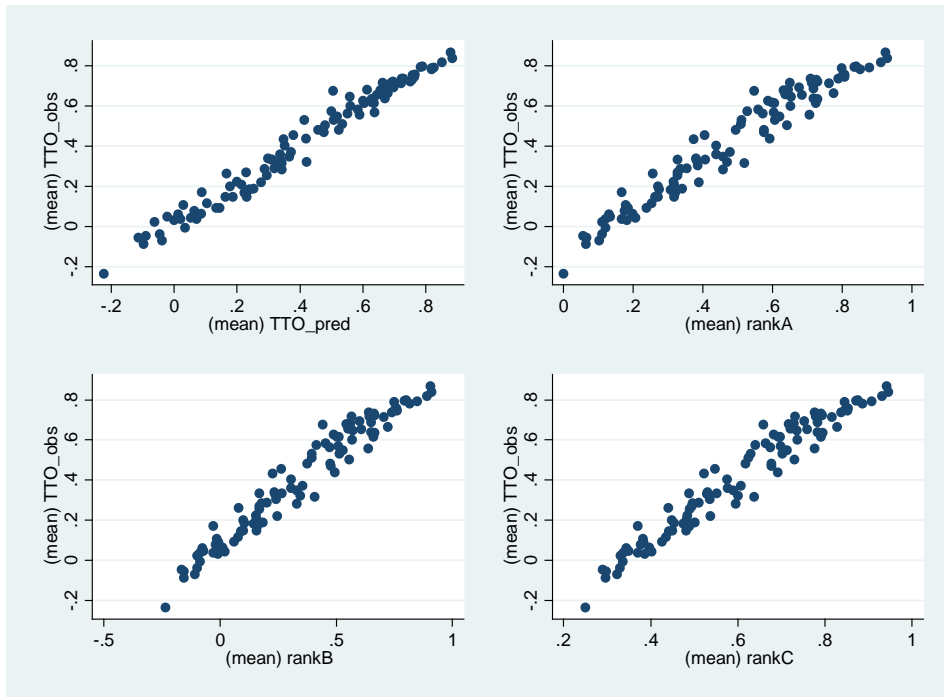
**FIGURA 8 – Valores observados da EAV e do TTO por nível de severidade do estado de saúde**



Fonte: elaboração do autor.

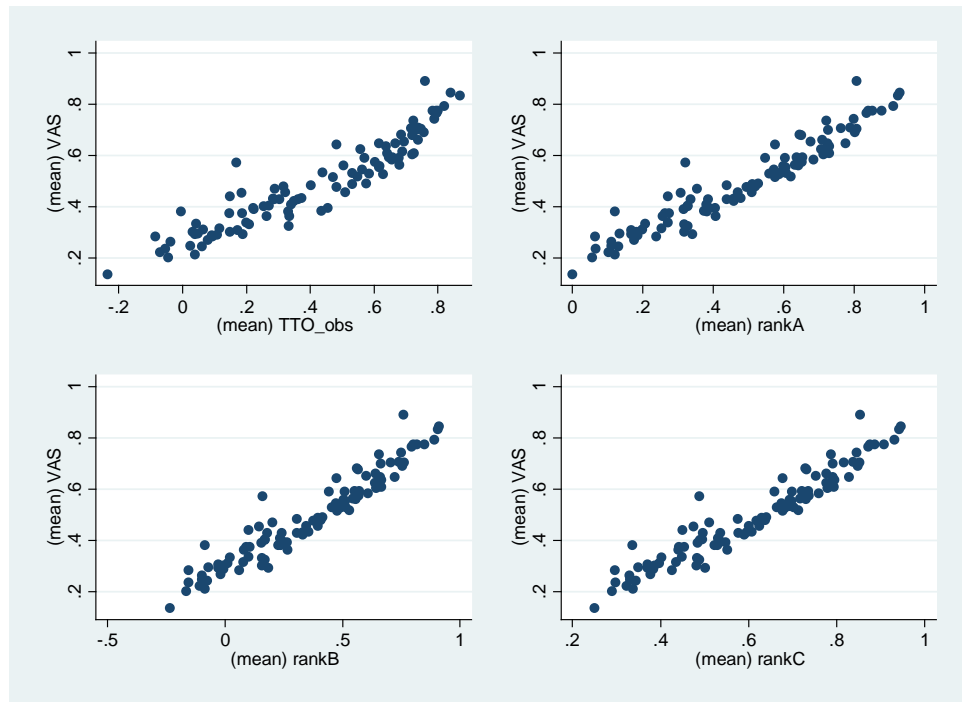
Nota: (1) estados brandos (2) estados moderados (3) estados severos.

**FIGURA 9 – Relação entre os pesos observados e preditos do TTO e do modelo logit condicional segundo três procedimentos de reescalonamento**



Fonte: elaboração do autor.

**FIGURA 10 – Relação entre os pesos observados da EAV e os valores preditos do modelo logit condicional segundo três procedimentos de reescalonamento**



Fonte: elaboração do autor.

## 5.2 Resultados das Inconsistências Lógicas

Os resultados encontrados para o caso brasileiro representam um baixo índice de inconsistências, já que apenas 8,7% dos pares de estados comparáveis no exercício TTO são inconsistentes, quando se considera o conceito de monotonicidade forte ou estrita das escolhas. A TABELA 9 apresenta as estatísticas sobre inconsistências lógicas para todos os métodos. Apesar de se observar uma distribuição aproximadamente equânime na proporção de indivíduos que cometem alguma inconsistência entre os métodos, a taxa média e mediana de inconsistência foi superior na métrica *choice-based*. Verifica-se que, enquanto no TTO, 8,7% de todos os pares de estados de saúde que possuem ordenamento lógico possível são inconsistentes, cerca de 6,0% e 4,9% dos pares que poderiam ser considerados inconsistentes são, de fato, classificados desta forma, no Exercício de Ordenação e na Escala Analógica Visual, respectivamente. Estes resultados indicam a existência de inconsistências secundárias, ou seja, aquelas derivadas, em grande parte, por algum aspecto do procedimento de medida.

Talvez, estes resultados evidenciem a dificuldade e o grau de abstração dos métodos baseados na troca, como TTO e SG. Nesse contexto, os indivíduos podem reportar valores de forma

aleatória, sem qualquer conformidade com os objetivos da pesquisa. Além disso, as métricas de ordenação e EAV possuem a vantagem dos respondentes poderem visualizar todos os estados de saúde simultaneamente, de forma a avaliar cada um deles com base não somente nos pontos de âncora, mas também com referência aos demais cartões. Este não é o caso do exercício da Troca de Tempo, no qual os únicos pontos de referência são a Morte e a Saúde Perfeita, a não ser que os indivíduos recordem todos os valores que atribuíram aos estados de saúde avaliados anteriormente, o que parece improvável.

**TABELA 9 – Inconsistências lógicas por métodos de avaliação**

	monotonicidade estrita			monotonicidade não-estrita		
	Ordenação	EAV	TTO	Ordenação	EAV	TTO
% respondentes inconsistentes	53,46	47,02	51,72	62,80	66,90	86,30
Taxa de Inconsistência Lógica						
Média	0,060	0,049	0,087	0,075	0,085	0,215
Mediana	0,045	0,000	0,067	0,050	0,053	0,182

Fonte: Elaboração do autor.

Quando se considera o conceito de monotonicidade fraca ou não-estrita, tanto a porcentagem de indivíduos inconsistentes, quanto as taxas médias e medianas de Inconsistência Lógica são menores para o Exercício de Ordenação se comparado aos demais métodos. Cerca de 86% dos respondentes produziram, no mínimo, uma inconsistência ao realizarem a tarefa do TTO, enquanto cerca de 63% e 67% dos indivíduos violaram a ordem lógica pré-estabelecida na Ordenação e na Escala Analógica Visual, respectivamente. Além disso, existe um maior número de respondentes, no método da Troca de Tempo, com taxa de inconsistência superior a 50% - 228 comparado a 7 e a 17 na ordenação e na EAV, respectivamente. Taxas de inconsistências médias são maiores que as medianas, sugerindo que alguns indivíduos altamente inconsistentes estão enviesando a média para cima.

Quanto à taxa de inconsistência, verifica-se que, enquanto no TTO, 21,5% de todos os pares de estados de saúde que possuem ordenamento lógico possível são inconsistentes, cerca de 7,5% e 8,5% dos pares que poderiam ser considerados inconsistentes são, de fato, classificados desta forma, no Exercício de Ordenação e na Escala Analógica Visual, respectivamente. Ser inconsistente na realização de uma das tarefas parece aumentar as chances de assim também o ser nos demais métodos. Os coeficientes de *Pearson* das taxas médias de inconsistências entre os instrumentos foram positivos a 1% de significância.

No caso da monotonicidade estrita, no qual não se consideram inconsistentes os indivíduos que oferecem mesmo valor ou posição a estados de saúde comparáveis diferentes, a dominância de inconsistências entre métodos se altera, com a Escala Analógica Visual apresentando um menor nível de violações. Estas diferenças evidenciam a frequência de “coincidências” entre os valores obtidos nos métodos cardinais vis-à-vis a ordenação. O TTO e a EAV, por serem medidas cardinais, sugerem um julgamento mais refinado e, portanto, uma maior diferenciação entre estados de saúde com descrições similares. Contudo, verifica-se a situação oposta, em que os instrumentos cardinais apresentam maiores frequências de valores idênticos para situações de saúde distintas.

Uma possível explicação para estes resultados é a engenhosidade da métrica TTO e, em menor medida, da EAV, e as consequências advindas desta complexidade no que se refere à compreensão dos exercícios por parte dos respondentes, ao alto grau de abstração necessário para a sua completude e à perda progressiva de estímulo resultante da morosidade do instrumento. Estes fatos podem contribuir para a maior indiferença dos respondentes entre os estados de saúde passíveis de comparação. Indivíduos desmotivados pela dificuldade no entendimento dos instrumentos ou cansados pelo relativo extenso período de duração da entrevista podem atribuir valores idênticos à distintas situações, de forma a demonstrar desinteresse e ansiedade pelo término das tarefas.

A existência de respondentes não-trocadores no Exercício de Troca do Tempo, ou seja, indivíduos que se recusam a trocar qualquer tempo de vida por uma melhora na qualidade de vida, atribuem o valor máximo aos pesos, independentemente do nível de severidade dos estados de saúde. Além disso, existe a possibilidade de alguns respondentes sabotadores, entendendo a sequência lógica do jogo, atribuírem um mesmo valor nas avaliações a fim de se atingir o ponto de indiferença precocemente, representando outra possibilidade para a evidencia de valorações iguais para condições de saúde distintas. A TABELA 10 revela o percentual de respondentes por número de inconsistências apresentadas e sua distribuição acumulada. Percebe-se que as valorações que violam a ordem lógica não representam as exceções. Quando não se considera inconsistência avaliar distintos estados de saúde com pesos iguais, o número máximo de inconsistências ocorre na EAV, que apresenta até 13 pares de estados comparáveis inconsistentes. No entanto, mais que 90% da amostra cometem até três inconsistências em todos os métodos. No caso da monotonicidade fraca ou não-estrita, apenas cerca de 1/3 da amostra é totalmente consistente em suas escolhas nos Exercícios de

Ordenação e Escala Analógica Visual, enquanto este percentual chega a menos que 15%, no caso do método da Troca de Tempo. Os números máximos de inconsistências crescem de forma expressiva na EAV e no TTO, chegando a 20 e a 15 pares, respectivamente.

**TABELA 10 – Percentual de respondentes inconsistentes (%) e distribuição acumulada, por método**

Pares Inconsistentes	monotonicidade estrita						monotonicidade não-estrita						
	Ordenação		EAV		TTO		Ordenação		EAV		TTO		
	(%)	F	(%)	F	(%)	F	(%)	F	(%)	F	(%)	F	
0	46,54	46,54	52,98	52,98	48,31	48,31	37,23	37,23	33,01	33,01	13,72	13,72	
1	23,32	69,86	22,96	75,94	23,21	71,52	25,23	62,47	26,04	59,05	21,44	35,16	
2	13,89	83,75	11,52	87,46	13,57	85,09	15,77	78,24	16,04	75,10	19,77	54,93	
3	7,03	90,78	5,48	92,94	7,67	92,75	9,22	87,46	10,21	85,30	16,56	71,49	
4	4,04	94,82	3,11	96,05	3,05	95,81	5,39	92,85	5,81	91,11	11,17	82,66	
5	2,16	96,98	1,74	97,79	1,86	97,66	3,02	95,87	3,98	95,09	5,78	88,44	
6	1,38	98,35	1,02	98,80	1,32	98,98	1,98	97,84	1,92	97,01	4,22	92,66	
7	0,81	99,16	0,60	99,40	0,51	99,49	1,11	98,95	1,02	98,02	2,73	95,39	
8	0,24	99,40	0,18	99,58	0,30	99,79	0,36	99,31	0,69	98,71	1,35	96,74	
9	0,30	99,70	0,09	99,67	0,15	99,94	0,30	99,61	0,42	99,13	0,90	97,63	
10	0,15	99,85	0,15	99,82	0,06	100,00	0,24	99,85	0,30	99,43	0,57	98,20	
11	0,06	99,91	0,12	99,94			0,06	99,91	0,21	99,64	0,45	98,65	
12	0,09	100,00	0,03	99,97			0,09	100,00	0,09	99,73	0,30	98,95	
13			0,03	100,00						0,15	99,88	0,60	99,55
14										.	.	0,09	99,64
15										0,03	99,91	0,36	100,00
16										0,03	99,94		
19										0,03	99,97		
20										0,03	100,00		

Fonte: Elaboração do autor.

A TABELA 11 apresenta a taxa média de inconsistência lógica para cada método, de acordo com as características socioeconômicas e demográficas dos respondentes. Um maior número de anos de educação formal e ser jovem parecem estar negativamente associados ao aumento das inconsistências, sendo que estas relações apresentam-se de forma quase que monotônica para o método da Troca de Tempo, considerando monotonicidade não-estrita. As variáveis referentes a condição socioeconômica e a posse de plano de saúde também estão associadas, de forma monotônica e significativa, ao número de inconsistências. Se o indivíduo não possui plano de saúde ou pertence às classes socioeconômicas mais empobrecidas, a taxa de inconsistência é maior. Pessoas viúvas, que se autoavaliam infelizes ou não muito felizes e que classificam a própria saúde como regular, ruim ou muito ruim apresentam maiores taxas médias de inconsistências. Sexo parece não estar associado ao nível de pares logicamente reversos na ordenação.

**TABELA 11 – Taxa média de inconsistência lógica por características sociodemográficas e de saúde**

Características Individuais	monotonicidade estrita			monotonicidade não-estrta		
	Ordenação	EAV	TTO	Ordenação	EAV	TTO
<b>Educação</b>						
Primário incompleto	0.071	0.045	0.111	0.083	0.080	0.252
Primário completo	0.065	0.058	0.095	0.080	0.094	0.238
Fundamental completo	0.069	0.055	0.100	0.086	0.089	0.224
Médio completo	0.055	0.044	0.076	0.069	0.079	0.203
Superior completo	0.038	0.039	0.058	0.047	0.070	0.161
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Faixa etária</b>						
18-25	0.060	0.047	0.089	0.072	0.076	0.199
26-35	0.060	0.053	0.087	0.078	0.094	0.212
36-45	0.064	0.053	0.084	0.079	0.090	0.224
46-55	0.061	0.047	0.089	0.075	0.081	0.229
55-64	0.053	0.047	0.083	0.066	0.080	0.221
	$p < 0.001$	$p < 0.001$	$p = 0.012$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Classe socioeconômica</b>						
A	0.035	0.025	0.069	0.049	0.061	0.185
B	0.052	0.044	0.075	0.067	0.077	0.199
C	0.066	0.055	0.093	0.080	0.090	0.221
D-E	0.080	0.060	0.124	0.092	0.097	0.281
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Possui Plano de Saúde</b>						
Sim	0.051	0.044	0.078	0.066	0.079	0.195
Não	0.064	0.052	0.091	0.079	0.087	0.225
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Sexo</b>						
Homem	0.059	0.049	0.087	0.074	0.086	0.219
Mulher	0.061	0.051	0.086	0.075	0.083	0.209
	$p = 0.257$	$p = 0.059$	$p = 0.337$	$p = 0.086$	$p = 0.011$	$p < 0.001$
<b>Estado Civil</b>						
Solteiro	0.060	0.047	0.090	0.075	0.080	0.210
Casado	0.058	0.051	0.084	0.073	0.087	0.215
Divorciado	0.065	0.051	0.075	0.078	0.089	0.218
Viúvo	0.073	0.045	0.129	0.087	0.083	0.268
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Saúde auto-avaliada</b>						
Bom ou Muito bom	0.057	0.046	0.085	0.073	0.082	0.208
Regular	0.069	0.062	0.095	0.081	0.095	0.239
Ruim ou Muito ruim	0.094	0.065	0.098	0.103	0.107	0.233
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<b>Felicidade</b>						
Feliz ou Muito feliz	0.058	0.048	0.086	0.072	0.083	0.214
Infeliz ou Não muito feliz	0.074	0.060	0.091	0.091	0.097	0.222
	$p < 0.001$	$p < 0.001$	$p = 0.017$	$p < 0.001$	$p < 0.001$	$p = 0.792$

Fonte: elaboração do autor.

Nota: Os p-valores apresentados foram estimados utilizando os teste Kruskal-Wallis e Mann-Whitney.

O teste t oferece resultados semelhantes.

Portanto, há evidências da existência de inconsistências primárias no estudo, ou seja, àquelas intimamente ligadas às limitações ou características individuais. Diante das assimetrias das

distribuições das taxas de inconsistências lógicas, evidenciadas através dos testes Jarque-Bera e Shapiro-Wilk ( $p < 0.01$ ), além do teste t, testes não-paramétricos (Kruskal-Wallis e Mann-Whitney) foram utilizados para testar as diferenças nas taxas de inconsistências.

Através do modelo logit, que possui como variável resposta a dummy que toma valor igual a 1 se o indivíduo avalia pelo menos um par de estados de saúde comparáveis de forma reversa à ordem lógica pré-estabelecida; e 0, se o indivíduo é plenamente consistente em suas decisões; avaliam-se os efeitos isolados das características individuais dos respondentes e do tempo de duração da entrevista sobre a probabilidade de apresentar inconsistências lógicas. Os casos de referências são: ser mulher, pertencer a uma classe econômica elevada, estar casado, possuir idade entre 18 e 25 anos, possuir ensino primário incompleto, não apresentar dificuldades na execução da tarefa da Troca de Tempo, não possuir experiência de trabalho na área da saúde, não possuir plano de saúde, avaliar a própria saúde como boa ou muito boa e se considerar infeliz ou pouco feliz.

Os resultados apresentados na TABELA 12 corroboram as evidências que, de fato, o nível educacional, a condição socioeconômica e a idade são determinantes importantes na geração de pares inconsistentes. Contudo, a importância destes elementos varia de acordo com a métrica utilizada na obtenção das preferências. Escolaridade possui efeitos negativos sobre inconsistências lógicas, tanto na ordenação quanto no método da Troca de Tempo, ao passo que na EAV, este efeito é positivo. Enquanto a diminuição das inconsistências só se faz sentir a partir do ensino médio e superior para o exercício de Ordenação, esta relação é extremamente forte no método da Troca de Tempo, de modo que qualquer ascensão no grau de escolaridade possui reflexos significativos na queda da inconsistência. Indivíduos com ensino superior completo possuem aproximadamente 60% a menos de chances de cometerem inconsistência, relativo aos respondentes que não completaram o ensino primário.

A idade parece possuir algum efeito somente no TTO. Contudo, a magnitude da diminuição das chances de produção de inconsistências é bastante irrisória. No que se refere ao nível socioeconômico, quanto pior as condições materiais do domicílio, maiores as chances de se cometer ordenações de forma a violar a ordem lógica, em todos os métodos. Os indivíduos cujo domicílio se encaixa nas classes D ou E possuem até 141% a mais de chances de cometer inconsistências que indivíduos residentes de domicílios mais abastados.

**TABELA 12 – Razão de chances das inconsistências lógicas por características sociodemográficas e de saúde**

Características Individuais	monotonicidade estrita					
	Ordenação		EAV		TTO	
	Odds	dp	Odds	dp	Odds	dp
Educação						
Primário completo	1,019	0,066	1,521***	0,098	0,678***	0,045
Fundamental completo	1,054	0,069	1,253***	0,082	0,754***	0,051
Médio completo	0,891*	0,059	1,126*	0,075	0,539***	0,037
Superior completo	0,657***	0,053	1,212**	0,097	0,432***	0,035
Idade	0,999	0,001	0,998	0,001	0,994***	0,001
Classe socioeconômica						
B	1,509***	0,095	1,669***	0,109	0,833***	0,051
C	1,672***	0,109	2,071***	0,139	0,909	0,058
D-E	2,413***	0,202	2,162***	0,182	1,178**	0,097
Homem	1,092***	0,027	1,004	0,024	0,955*	0,023
Estado Civil						
Viúvo	1,034	0,083	0,858*	0,068	1,423***	0,116
Divorciado/Separado	0,979	0,046	0,896**	0,042	0,832***	0,039
Solteiro	0,970	0,028	0,948*	0,027	1,103***	0,031
Trabalhou na area da saúde	0,875***	0,032	0,871***	0,032	1,156***	0,043
Possui Plano de Saúde	0,999	0,027	0,997	0,027	0,985	0,026
Saúde auto-avaliada						
Regular	1,024	0,033	1,229***	0,039	1,064*	0,035
Ruim ou Muito ruim	0,983	0,091	0,844*	0,077	1,137	0,104
Feliz	0,775***	0,029	0,811***	0,031	0,900***	0,034
Tempo de Entrevista	1,000	0,000	1,001***	0,000	1,000	0,000
Dificuldade no TTO	1,104***	0,027	0,841***	0,020	0,836***	0,020
Constante	0,873	0,099	0,535***	0,061	2,730***	0,310

Fonte: Elaboração do autor.

\*\*\* 1% de significância \*\* 5% de significância \* 10% de significância

Quanto às demais variáveis, um resultado contra-intuitivo se faz presente: apresentar dificuldades na execução das tarefas do *TTO* diminui a probabilidade de ser inconsistente nos métodos cardinais. Talvez, este resultado possa ser explicado pelo alto nível de esforço e empenho que os indivíduos impuseram no desenvolvimento destas tarefas a fim de compreender a natureza do jogo, diferentemente dos respondentes desinteressados ou que não entenderam o objetivo dos instrumentos e que, portanto, resultaram em maiores números de inconsistências. Tempo de duração da entrevista e possuir plano de saúde parecem não apresentar qualquer efeito.

Sexo possui efeitos de pequena magnitude, mas significativos. Enquanto homens apresentam 9,2% a mais de chances de cometerem inconsistências na ordenação, as mulheres possuem quase 5% a mais de chances de gerarem ordenações reversas no *TTO*. Ter experiência de

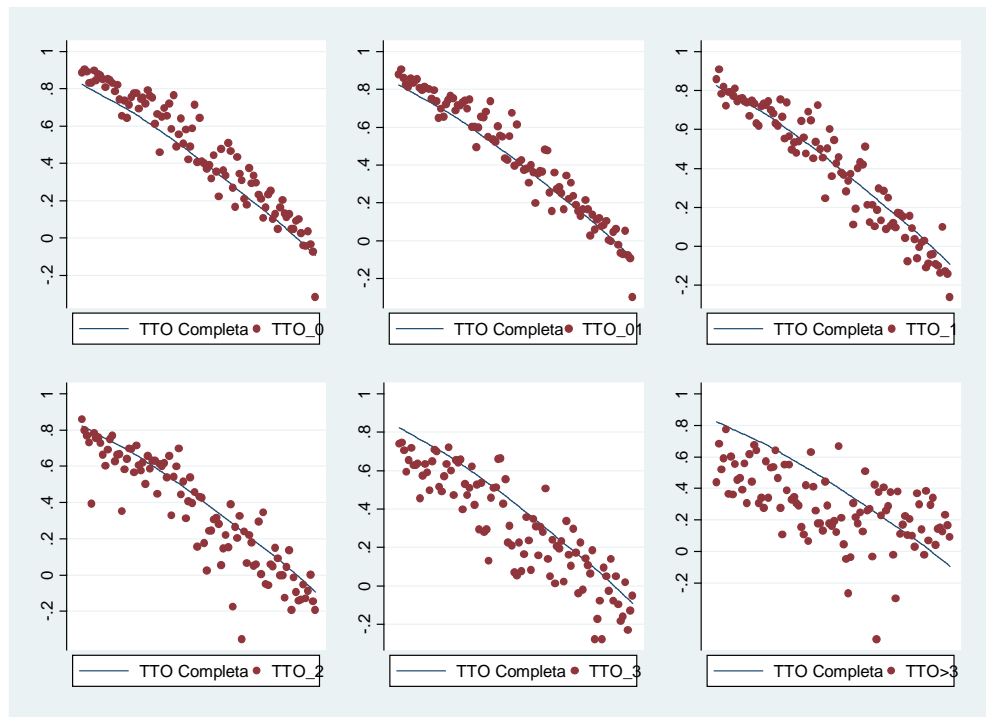
trabalho na área da saúde e se perceber feliz ou muito feliz diminuem as chances de se cometer inconsistência, principalmente, na Ordenação e na EAV. Saúde auto-avaliada possui efeitos somente na Escala e os viúvos possuem cerca de 42% a mais de chances de cometer julgamentos reversos à ordem lógica pré-estabelecida no método TTO, em relação aos casados.

Quando se considera o conceito de monotonicidade não-estrita, os efeitos da saúde auto-avaliada sobre as inconsistências passam a ser significativos, com os indivíduos que dizem possuir saúde regular, ruim ou muito ruim apresentando uma redução nas chances de se cometer inversões na ordenação, em relação aos respondentes que percebem sua saúde como boa ou muito boa. No que se refere à métrica TTO, possuir plano de saúde passa a possuir efeito significativo na redução das chances de se cometer inconsistências, enquanto ser feliz apresenta efeito oposto. Quando as avaliações são realizadas a partir da EAV, o aumento da idade eleva, de forma significativa, esta chance.

Estes resultados mostram evidências da presença de inconsistências primárias, ou seja, àquelas geradas por limitações dos respondentes. Sendo assim, se os julgamentos reversos são mais frequentes nos subgrupos populacionais de menor escolaridade ou em pior situação socioeconômica, a exclusão de respostas inconsistentes do cálculo final das tarifas pode levar, no limite, à geração de médias que sub-representam os gostos destes grupos populacionais. Algumas diferenças são percebidas, no que se refere aos efeitos das inconsistências lógicas sobre os pesos dos estados de saúde. As figuras 11, 12, 13, 14 e 15 apresentam as comparações dos pesos médios obtidos de subamostras com diferentes níveis de inconsistência dos respondentes, tendo como base a amostra com todos os indivíduos participantes.

Considerando-se o conceito de monotonicidade estrita das escolhas na definição de inconsistência, as subamostras com respondentes consistentes (TTO\_0) ou com, no máximo, uma inconsistência (TTO\_01), apresentam pesos médios observados do TTO superiores àqueles encontrados para a amostra completa, principalmente para os estados de saúde brandos e moderados. Já na EAV, os valores médios observados das referidas subamostras apresentam-se ligeiramente inferiores aos encontrados na amostra completa, principalmente para os estados severos.

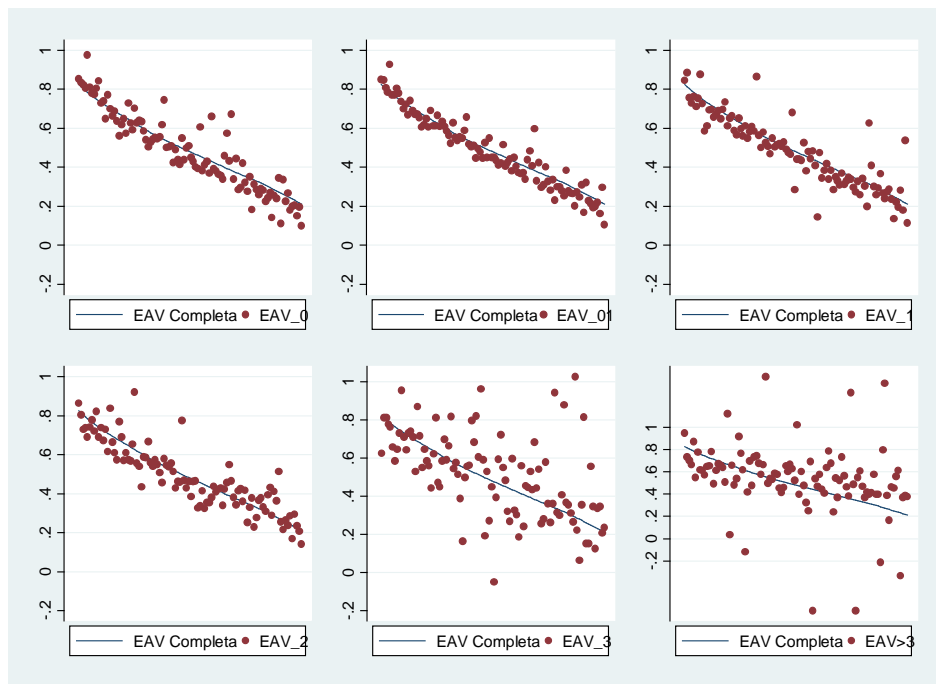
**FIGURA 11 – Valores médios observados do TTO segundo nível de inconsistências**



Fonte: Elaboração do autor.

Nota: TTO<sub>i</sub> refere-se à amostra com respondentes que apresentaram um número *i* de inconsistências. Assim, TTO<sub>0</sub>, TTO<sub>01</sub>, TTO<sub>2</sub> e TTO<sub>>3</sub> representam as amostras com 0, 0-1, 2 e mais que 3 inconsistências no TTO, por indivíduo.

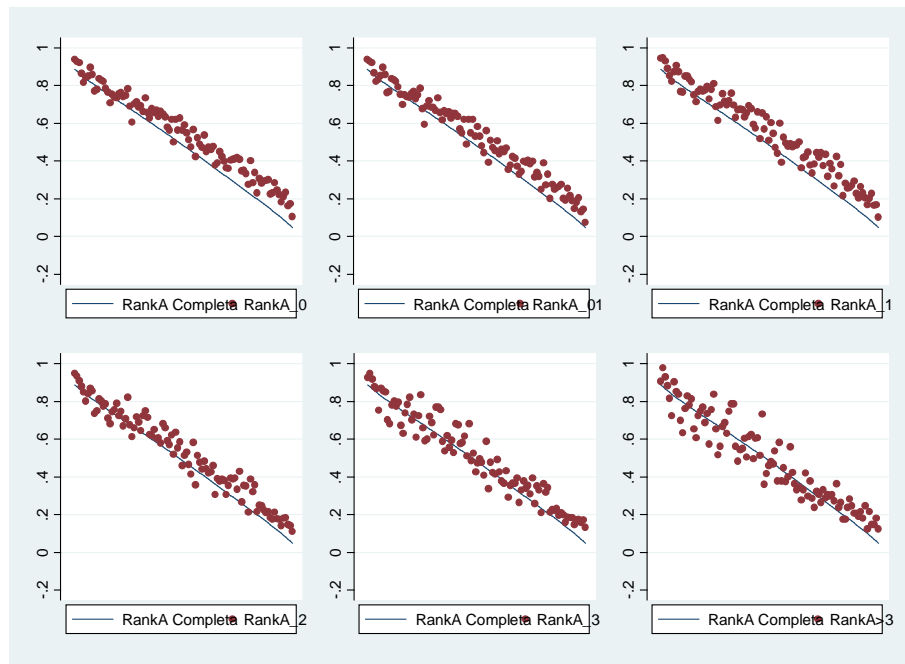
**FIGURA 12 – Valores médios observados da EAV segundo nível de inconsistências**



Fonte: Elaboração do autor.

Nota: EAV<sub>i</sub> refere-se à amostra com respondentes que apresentaram um número *i* de inconsistências. Assim, EAV<sub>0</sub>, EAV<sub>01</sub>, EAV<sub>2</sub> e EAV<sub>>3</sub> representam as amostras com 0, 0-1, 2 e mais que 3 inconsistências na EAV, por indivíduo.

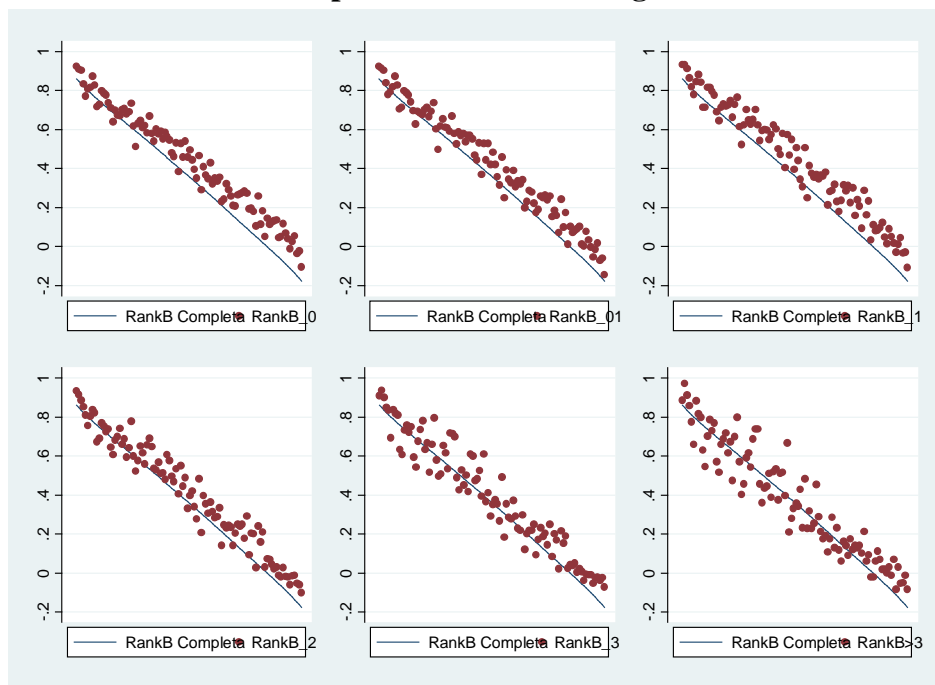
**FIGURA 13 – Valores médios preditos do Rank A segundo nível de inconsistências**



Fonte: Elaboração do autor.

Nota: RankA<sub>i</sub> refere-se à amostra com respondentes que apresentaram um número *i* de inconsistências. Assim, RankA\_0, RankA\_01, RankA\_2 e RankA\_>3 representam as amostras com 0, 0-1, 2 e mais que 3 inconsistências no Rank A, por indivíduo.

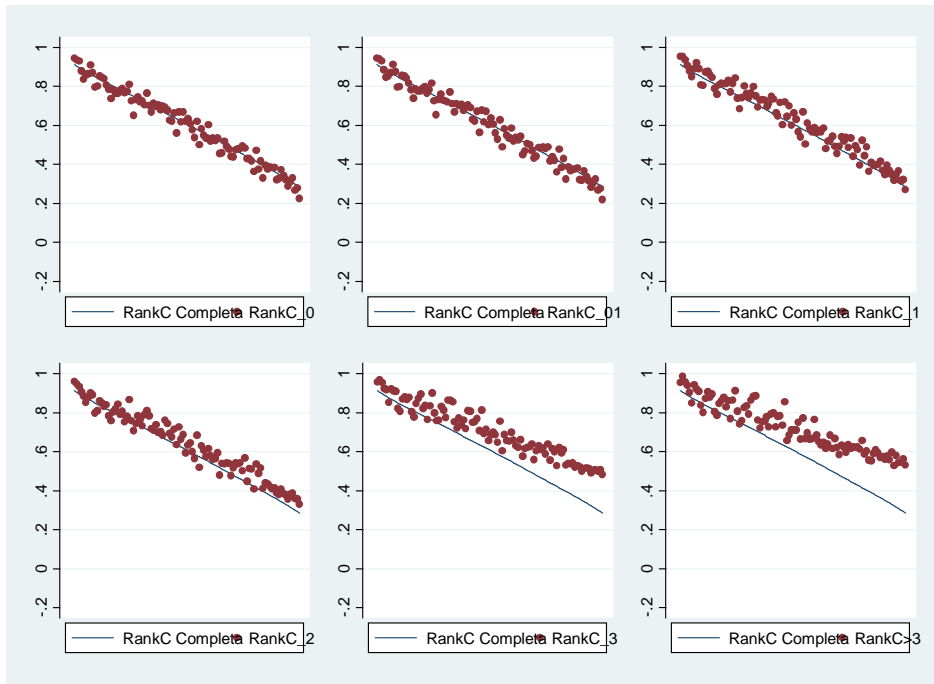
**FIGURA 14 – Valores médios preditos do Rank B segundo nível de inconsistências**



Fonte: Elaboração do autor.

Nota: RankB<sub>i</sub> refere-se à amostra com respondentes que apresentaram um número *i* de inconsistências. Assim, RankB\_0, RankB\_01, RankB\_2 e RankB\_>3 representam as amostras com 0, 0-1, 2 e mais que 3 inconsistências no Rank B, por indivíduo.

**FIGURA 15 – Valores médios preditos do Rank C segundo nível de inconsistências**



Fonte: Elaboração do autor.

Nota: RankC\_i refere-se à amostra com respondentes que apresentaram um número i de inconsistências. Assim, RankC\_0, RankC\_01, RankC\_2 e RankC>3 representam as amostras com 0, 0-1, 2 e mais que 3 inconsistências no Rank C, por indivíduo.

Quando se consideram os pesos preditos para o Exercício de Ordenação, os indivíduos consistentes apresentam maiores médias que aqueles menos consistentes, tanto no Rank A quanto no Rank B. Considerando-se os diferentes níveis de inconsistências lógicas definidas pelo EQ-5D, observa-se maior dispersão dos pesos médios em relação à amostra completa, à medida que o número de inconsistências aumenta. Os valores mais errantes são advindos das subamostras compostas por indivíduos com três ou mais inconsistências em suas avaliações. Isso ocorre tanto no método da Troca de Tempo quanto na Escala Analógica Visual, no qual também se verifica queda nos valores para estados de saúde brandos e aumento dos valores para estados severos, à medida que se transita de subamostras consistentes para subamostras de indivíduos com maior nível de inconsistências.

No que se refere aos pesos preditos para o Exercício de Ordenação, o Rank A e o Rank B apresentam maior estabilidade das médias ao longo das subamostras. A forma de reescalonamento que atribui utilidade zero ao estado Morte - Rank C - apresenta o pior ajuste, com os valores médios para estados moderados e severos superiores aos correspondentes valores associados as subamostras mais consistentes. Quando se considera como

inconsistentes os indivíduos que, em um determinado par de estados de saúde, atribuem valores maiores ou iguais a estados dominados, os pesos médios se comportam de forma semelhante em todas as subamostras analisadas, apresentando-se, porém, em um nível inferior aos encontrados para o caso de monotonicidade estrita das escolhas.

De acordo com a TABELA 13, em todos os métodos, a subamostra de indivíduos com, no máximo, uma inconsistência, apresentou o conjunto de pesos mais similares ao da amostra completa. O coeficiente de *Pearson* foi de 0.989 para os pesos obtidos pelo método EAV da amostra completa e da subamostra de respondentes com zero ou uma inconsistência. Já para amostra de indivíduos que apresentaram mais que três inconsistências na escala, a correlação com a amostra completa foi de apenas 0.367. No que se refere aos dados derivados a partir da Ordenação, para a referida subamostra, as correlações são quase perfeitas.

**TABELA 13 – Correlação de Pearson entre os pesos da amostra completa e das subamostras definidas a partir do número de inconsistências**

	<b>TTO</b>	<b>EAV</b>	<b>Rank A</b>	<b>Rank B</b>	<b>Rank C</b>
0	0,987	0,964	0,998	0,998	0,998
0-1	0,995	0,989	0,999	0,999	0,999
1	0,979	0,886	0,998	0,998	0,998
2	0,935	0,905	0,995	0,995	0,996
3	0,899	0,550	0,985	0,985	0,985
> 3	0,574	0,367	0,968	0,968	0,967

Fonte: Elaboração do autor.

Nota: considera-se monotonicidade estrita na definição de inconsistências. Os coeficientes de *Pearson* para o caso de monotonicidade fraca ou não-estrita das escolhas seguem padrões semelhantes, apresentando, contudo, menores níveis de correlação.

Se considerarmos a censura de dados inconsistentes, comumente realizada na literatura, os achados encontrados neste trabalho revelam que, para o caso de Minas Gerais, a exclusão destas informações afetaria muito sutilmente os pesos populacionais dos estados de saúde. Talvez, este achado seja resultante da grande proporção de indivíduos consistentes ou com apenas uma inconsistência na amostra completa. Acredita-se que o grande esforço empreendido pelas equipes envolvidas na fase de coleta dos dados foi importante para a qualidade alcançada na amostra, principalmente no que tange ao monitoramento dos pesquisadores quanto à condução das entrevistas e ao entendimento dos instrumentos de avaliação.

A fim de se observar o efeito da exclusão de indivíduos inconsistentes da amostra sobre os coeficientes utilizados para a interpolação de estados de saúde que não foram observados diretamente e as medidas de ajustes das métricas cardinais e ordinais, a TABELA 14 apresenta os resultados do modelo logit condicional, segundo o reescalonamento de melhor ajuste - Rank B, e os resultados do modelo de efeitos aleatórios para o método da Troca de Tempo, utilizando tanto a amostra completa quanto a amostra censurada, que exclui os indivíduos com mais de uma inconsistência em suas avaliações.

**TABELA 14 – Coeficientes estimados a partir do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo) para as amostras completa e consistente**

Variáveis	Coeficientes TTO	Coeficientes TTO_01	Coeficientes Rank	Coeficientes Rank_01
Mobilidade 2	0,128***	0,109***	1,143***	1,429***
Cuidados Pessoais 2	0,121***	0,122***	0,689***	0,859***
Atividades Habituais 2	0,095***	0,088***	0,503***	0,658***
Dor/Desconforto 2	0,067***	0,082***	0,436***	0,583***
Ansiedade/Depressão 2	0,062***	0,070***	0,405***	0,523***
Mobilidade 3	0,404***	0,404***	2,307***	3,107***
Cuidados Pessoais 3	0,247***	0,271***	1,052***	1,443***
Atividades Habituais 3	0,205***	0,228***	0,869***	1,286***
Dor/Desconforto 3	0,200***	0,244***	0,752***	1,078***
Ansiedade/Depressão 3	0,113***	0,135***	0,662***	0,922***
Diferença Absoluta Média	0,035	0,038	0,073	0,081
No. >0.05	25	29	64	64
No. >0.10	2	4	29	38
rho de Pearson	0,987	0,987	0,974	0,973
rho de Spearman	0,987	0,988	0,973	0,975

Fonte: elaboração do autor.

\*\*\* Significante a 1%.

No caso do TTO, a maior diferença entre os coeficientes se deu na presença de alguns problemas na dimensão de mobilidade. Os coeficientes apresentaram pequenas diferenças ao se excluir indivíduos com maior número de inconsistências. Contudo, a importância relativa das dimensões se alterou: problemas moderados de mobilidade passam a gerar menor desutilidade que problemas com os cuidados pessoais, e problemas extremos com as atividades habituais geram menores prejuízos para a qualidade de vida que dores ou desconforto extremos. As diferenças absolutas médias e as correlações entre os pesos sofreram aumentos sutis em seus valores, mas elevaram-se os números de estados de saúde cujas diferenças absolutas entre pesos foram superiores a 0.05 e 0.10.

Quanto aos valores estimados com base nos dados ordenados, apesar dos coeficientes apresentarem maiores diferenças ao se excluir indivíduos com maior número de inconsistências, a importância relativa das dimensões permaneceu inalterada. Os problemas

severos na dimensão de mobilidade continuam gerando os maiores decréscimos de utilidade, sendo que os coeficientes associados a estes problemas apresentam a maior diferença entre a amostra completa e a subamostra consistente. Em ambos os grupos, depois das complicações na dimensão de mobilidade, problemas ligados aos cuidados pessoais possuem maior importância que dificuldades na realização das atividades habituais, seguidas por dor ou desconforto e, finalmente, ansiedade ou depressão. Apesar da correlação linear quase não se alterar, as estatísticas de qualidade de ajuste apresentam maior variação, mantendo-se a diferença absoluta média inferior a 10%.

## 6. CONSIDERAÇÕES FINAIS

Este trabalho tem como objetivos verificar a viabilidade do uso da abordagem ordinal em análises de custo-utilidade e analisar em que medida três diferentes métodos de obtenção de preferências populacionais referentes à saúde – Exercício de Ordenação, Escala Analógica Visual e Time Trade-Off - são equivalentes, tanto do ponto de vista das utilidades associadas aos estados de saúde, quanto no que se refere ao grau de consistência que apresentam.

O Time Trade-Off é o método dominante na avaliação de estados de saúde. Contudo, tem sido criticado pelo maior peso cognitivo e pela dificuldade de administração, pela hipótese de proporcionalidade constante e pela transformação dos valores dos estados classificados como piores que a morte (CRAIG, BUSSCHBACH e SALOMON, 2009b). Os resultados apresentados nesta dissertação suportam o uso de um exercício de avaliação mais simples. A utilização de dados ordenados pode possibilitar a incorporação das preferências de grupos populacionais para a qual o TTO é bastante árduo, como os menos escolarizados e aqueles pertencentes às classes sociais mais desfavorecidas. Além disso, esta métrica facilita a operacionalização do estudo ao diminuir o requerimento de vastos recursos para a pesquisa de campo, diminuir o tempo de duração de cada entrevista e diminuir o erro sistemático do entrevistador.

Dada a diferença da natureza das informações sob as quais os modelos são baseados, o modelo logit condicional apresenta uma performance de ajuste aceitável, com diferença absoluta média inferior a 10% para o Rank A e o Rank B. Todos os valores preditos são altamente correlacionados, com grande parte dos coeficientes de correlação acima de 0.95. Entretanto, ao mesmo tempo em que as estratégias ordinais não são contaminadas por questões relacionadas à duração ou a transformações referentes aos estados piores que a morte, estas métricas contêm seus próprios problemas, que necessitam ser explorados no contexto de avaliação de estados de saúde. Todos os modelos obtiveram maiores erros médios absolutos para estados de saúde severos. Este fato, supostamente relaciona-se à maior facilidade com que o método TTO possibilita a troca entre tempo e qualidade de vida, registrando menores valores para estados severos, quando comparado às demais métricas de avaliação (CRAIG *et al.*, 2009).

A maior dificuldade dos indivíduos em classificar estados de saúde como piores que a morte na métrica ordinal, também imprime dificuldades na estimativa do coeficiente para o estado Morte, no Rank C, necessária para normalizar os pesos na escala 0-1. Se as escolhas envolvendo a morte são determinísticas, contraria-se o pressuposto de aleatoriedade do modelo de utilidade aleatória, sob o qual baseiam-se os modelos logit condicionais (FLYNN *et al.*, 2008). Este fato pode explicar parte das diferenças observadas entre os resultados das diferentes formas de reescalonamento. Assim, maior importância deve ser dada às estimações de resultados cardinais a partir de informações ordinais, dado que sua utilização em análises de custo-utilidade ainda apresenta desafios quanto à forma de normalização. Além disso, pesquisas sobre o processo de reflexão dos indivíduos ao realizarem a tarefa de ordenação poderiam contribuir de forma expressiva para se avaliar a consistência do processo de tomada de decisões dos indivíduos e as hipóteses dos modelos estatísticos.

Quanto aos métodos cardinais, os resultados da Escala Analógica Visual e do TTO apresentaram a menor correlação linear. A curvatura na relação entre os valores observados sugere uma forma funcional quadrática para a função valor, geralmente estimada a fim de se transformar pesos obtidos através do método EAV em utilidades TTO/SG (TORRANCE, 1976; ROBINSON, LOOMES e JONES-LEE, 2001). A relação encontrada confirma os achados por *Craig et al.* (2009a) para os países europeus. Já no que se refere à qualidade das informações, os resultados encontrados para o caso brasileiro apresentam um baixo índice de inconsistências, dado que apenas 8,7% dos pares de estados comparáveis no exercício TTO são inconsistentes, quando se considera o conceito de monotonicidade forte ou estrita das preferências. Talvez, a alta consistência dos dados seja um reflexo do esforço empreendido, ainda na fase de pesquisa de campo, no acompanhamento direto por parte dos pesquisadores sobre a qualidade de aplicação dos questionários e sobre a conduta dos entrevistadores.

Apesar da proporção de respondentes inconsistentes ser similar aquela achada na literatura internacional, as taxas de inconsistência foram significativamente superiores às encontradas por Badia, Monserrat e Herdman (1999), em estudo para a Espanha que também utilizou o sistema descritivo EQ-5D, indicando a possibilidade de maior ocorrência de inconsistências no contexto de populações mais heterogêneas do ponto de vista socioeconômico (CRAIG e RAMACHANDRAN, 2006). O Exercício de Troca de Tempo produziu maiores taxas de inconsistência que a Escala Analógica Visual e o Exercício de Ordenação, independentemente da definição adotada na construção da taxa. Este resultado indica a existência de

inconsistências secundárias, ou seja, aquelas derivadas, em grande parte, por algum aspecto do procedimento de medida. Esta evidência também deve ser considerada pelos pesquisadores no momento da escolha das métricas de mensuração de preferências.

Talvez, estes resultados evidenciem o grande peso cognitivo e o grau de abstração dos métodos baseados na troca, como TTO e SG. As diferenças entre resultados, considerando os dois conceitos de inconsistências, monotonicidade fraca e forte, evidenciam a frequência de “coincidências” entre os valores obtidos nos métodos cardinais vis-à-vis a ordenação. Verifica-se a situação oposta ao esperado, em que os instrumentos cardinais apresentam maiores frequências de valores idênticos para situações de saúde distintas. Os achados deste estudo corroboram aqueles encontrados por Craig *et al.* (2009a), identificando incongruências na própria estrutura teórica dos métodos cardinais.

O nível educacional e a condição socioeconômica são determinantes importantes na geração de pares inconsistentes. Contudo, a importância destes elementos varia de acordo com a métrica utilizada na obtenção das preferências. Enquanto a diminuição das inconsistências só se faz sentir a partir do ensino médio e superior para o exercício de Ordenação, esta relação é extremamente forte no método da Troca de Tempo, de modo que qualquer ascensão no grau de escolaridade possui reflexos significativos na queda da inconsistência nesta métrica. Talvez, este resultado reflète o alto nível de abstração envolvido no julgamento de estados de saúde hipotéticos, independentemente de qual método está sendo utilizado, e o alto peso cognitivo, quando se refere à métrica TTO. Apesar de significativo, o efeito da idade sobre a geração de inconsistências no método TTO é irrisório. Quanto pior as condições materiais do domicílio, maiores as chances de se cometer ordenações de forma a violar a ordem lógica, em todos os métodos. Estes resultados mostram evidências da presença de inconsistências primárias, ou seja, àquelas geradas por limitações dos respondentes.

Observa-se maior dispersão dos pesos médios em relação à amostra completa, à medida que o número de inconsistências aumenta. Também se verifica queda nos valores para estados de saúde brandos e aumento dos valores para estados severos, à medida que se transita de subamostras consistentes para subamostras de indivíduos com maior nível de inconsistências. No que se refere aos pesos preditos para o Exercício de Ordenação, o Rank A e o Rank B apresentam maior estabilidade das médias ao longo das subamostras. Em todos os métodos, a subamostra de indivíduos com, no máximo, uma inconsistência, apresentou o conjunto de

pesos mais similares ao da amostra completa. Se considerarmos a censura de dados inconsistentes, comumente realizada na literatura, os achados encontrados neste trabalho revelam que, para o caso de Minas Gerais, a exclusão destas informações afetaria muito sutilmente os pesos populacionais dos estados de saúde, em conformidade com os achados de Lamers *et al.* (2006), para o caso holandês, e de Devlin *et al.* (2003), para a Nova Zelândia.

Talvez, dois fatores tenham contribuído para este resultado. Primeiro, a grande proporção de indivíduos consistentes ou com apenas uma inconsistência na amostra completa. Acredita-se que o grande esforço empreendido pelas equipes envolvidas na fase de coleta dos dados foi importante para a qualidade alcançada na amostra, principalmente no que tange ao monitoramento dos pesquisadores quanto à condução das entrevistas e ao entendimento dos instrumentos de avaliação. Segundo, que a estrutura de preferência dos excluídos talvez não difere muito daquelas de outros subgrupos ou da média populacional. Existem evidências de que, para Minas Gerais, enquanto indivíduos de grupos etários diferentes possuem utilidades estatisticamente distintas entre si, as variáveis de escolaridade e condição socioeconômica não são significativas para explicar diferenças dos pesos médios (MESQUITA, 2013). Quando esta informação é aliada ao efeito ínfimo encontrado para a variável idade na explicação das inconsistências, pode-se inferir o baixo impacto da censura da amostra sobre os resultados encontrados.

Neste trabalho verificou-se o impacto das exclusões apenas nos pesos observados do TTO e da EAV. Para fins de comparação com a modelagem padrão, realizada para a interpolação dos 141 estados de saúde restantes do sistema descritivo EQ-5D (ANDRADE *et al.*, 2013), os coeficientes das variáveis de dimensão de saúde são estimados a partir da amostra consistente, para o método da Troca de Tempo. Verifica-se que a importância relativa das dimensões de saúde se altera entre os coeficientes estimados com a amostra completa e com a subamostra consistente. Assim, se na amostra completa, mobilidade e atividades usuais geravam, respectivamente, maiores desutilidades que cuidados pessoais e dor/desconforto; na amostra consistente, observa-se comportamento contrário. Contudo, as magnitudes dos coeficientes foram pouco alteradas. Uma agenda de pesquisa é analisar o efeito destas mudanças das utilidades relativas às dimensões de saúde, sobre os pesos médios estimados e, conseqüentemente, sobre a decisão de alocação dos recursos. Quanto aos valores estimados com base nos dados ordinais, apesar dos coeficientes apresentarem maiores diferenças ao se

excluir indivíduos com maior número de inconsistências, a importância relativa das dimensões permaneceu inalterada.

Diferentemente deste estudo, altas proporções de respondentes com muitas inconsistências em suas avaliações podem resultar em estimações enviesadas dos pesos populacionais. Sociedades com alto grau de heterogeneidade socioeconômica e cultural, associadas a diferenças intergrupos das estruturas de preferências ou a um baixo esforço de pesquisa podem apresentar maior sensibilidade das tarifas em relação à censura dos dados inconsistentes. Torna-se interessante pensar melhorias nas técnicas adotadas ou ampliar o escopo de métricas de obtenção de pesos através de instrumentos alternativos, a fim de se avançar em termos de qualidade dos dados e representatividade populacional. O menor nível de inconsistências encontrado no Exercício de Ordenação também sugere suas potencialidades em avaliações econômicas em saúde. O grande ajuste encontrado pelo exercício evidencia sua grande viabilidade, principalmente, em análises de tecnologias associadas a estados brandos e moderados de saúde.

Quando se tomam decisões sobre alocação de recursos entre programas de saúde alternativos, principalmente quando se utilizam pesquisas populacionais, a factibilidade pode ser um fator decisivo para a escolha do método. Assim, o fato de algumas técnicas não se basearem na teoria da decisão, não permite seu abandono ou subutilização. Muitas hipóteses podem ser satisfeitas de forma aproximada e o julgamento final sobre as métricas deve considerar a extensão em que a perda da racionalidade microeconômica é compensada por uma maior praticidade. Este estudo não resolve o intenso debate sobre qual método se mostra superior, mas evidencia diferenças importantes em aspectos específicos das técnicas. Acredita-se que este debate ganha ainda mais força quando se consideram contextos populacionais mais heterogêneos e com menor nível de desenvolvimento. A adoção de padrões internacionais sem um questionamento crítico com respeito às particularidades regionais pode acarretar sérias implicações para a qualidade e representatividade das informações e, por conseguinte, para as análises de custo-utilidade. Uma agenda de pesquisa é extrapolar os efeitos do uso de diferentes metodologias sobre os pesos populacionais, verificando quais seus impactos diretos na análise de custo-utilidade e nas políticas públicas que dela decorrer.

Por fim, apesar das análises de custo-utilidade em saúde auxiliarem o processo de tomada de decisão, outras considerações também devem ser incluídas neste processo. Muitas

intervenções tecnológicas apresentam efeitos muito mais amplos que as estratégias para as quais são direcionadas. Normalmente, este campo de ação do efeito não é incluído de forma explícita no processo de tomada de decisão. Equidade e aspectos distributivos também devem ser considerados. O uso restrito de listas de custo-efetividade pode resultar em anomalias na alocação de recursos, deixando, por exemplo, algumas doenças absolutamente sem tratamento em favor de outras. Doenças raras, mas potencialmente curáveis, podem ter um manejo muito desfavorável do ponto de vista econômico. Assim, questões que transcendem a eficiência econômica devem sempre ser avaliadas (BRASIL, 2008).

## REFERÊNCIAS

- ALI, S.; RONALDSON, S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British Medical Bulletin*, v. 103, p. 1-24, 2012.
- ANDRADE, M. *et al.* Societal Preferences for EQ-5D Health States from a Brazilian Population Survey. *Value in Health Regional Issues*, v. 2, n. 3, p. 405-412, Dec. 2013.
- ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA - ABEP. *Critério de classificação econômica Brasil*. Disponível em: <<http://www.abep.org/novo/Content.aspx?ContentID=835>>. Acesso em: 4 de fevereiro de 2013.
- AUGUSTOVSKI, F. *et al.* Argentine Valuation of the EQ-5D Health States. *Value In Health*, v. 12, n. 4, p. 587-596, 2009.
- BADIA, X. *et al.* Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and time trade-off. *Quality of Life Research*, v. 8, p. 303-310, 1999.
- BADIA, X.; MONSERRAT, R.; HERDMAN, M. Inconsistent responses in three preference-elicitation methods for health states. *Social Science & Medicine*, v. 49, n. 7, p. 943-950, Oct. 1999.
- BANSBACK, N. *et al.* Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. *PLoS ONE*, v. 7, n. 2, p. 1-11, Feb. 2012a.
- BANSBACK, N. *et al.* Using a Discrete Choice Experiment to Estimate Health State Utility Values. *Journal of Health Economics*, v. 31, n. 1, p. 306-318, Jan. 2012b.
- BEKKER-GROB, E.; RYAN, M.; GERARD, K. Discrete choice experiments in health economics: a review of the literature. *Health Economics*, v. 21, n. 2, p. 145-172, Feb. 2012.
- BERNET, S. *et al.* Comparison of Different Valuation Methods for Population Health Status Measured by the EQ-5D in Three European Countries. *Value in Health*, v. 12, n. 5, p. 750-758, Mar. 2009.
- BLEICHRODT, H.; JOHANNESSON, M. The validity of QALYs: an experimental test of constant proportional trade-off and utility independence. *Medical Decision Making*, v. 17, n. 1, p. 21-32, 1996.
- BLEICHRODT, H.; JOHANNESSON, M. An Experimental Test of a Theoretical Foundation for Rating-scale Valuations. *Medical Decision Making*, v. 17, n. 2, p. 208-216, Apr-Jun. 1997.
- BRASIL. Secretaria de Políticas de Saúde. Política Nacional de Medicamentos. *Revista de Saúde Pública*, São Paulo, v. 34, n. 2, p. 206-209. Apr. 2000.
- BRASIL. MINISTÉRIO DA SAÚDE. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Avaliação de tecnologias em saúde:

institucionalização das ações no Ministério da Saúde. *Revista de Saúde Pública*, Brasília, v. 40, n. 4, p. 743-747, 2006.

BRASIL. MINISTÉRIO DA SAÚDE. Secretaria-Executiva. Área de Economia da Saúde e Desenvolvimento. *Avaliação econômica em saúde: desafios para gestão no Sistema Único de Saúde*. Brasília: Editora do Ministério da Saúde, 2008. 104 p.

BRASIL. MINISTÉRIO DA SAÚDE. Secretaria-Executiva. Área de Economia da Saúde e Desenvolvimento. *Avaliação de tecnologias em saúde: ferramentas para a gestão do SUS*. Brasília: Editora do Ministério da Saúde, 2009a. 110 p.

BRASIL. MINISTÉRIO DA SAÚDE. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. *Diretrizes Metodológicas: estudos de avaliação econômica de tecnologias em saúde*. Brasília: Ministério da Saúde, 2009b. 150 p.

BRASIL. Lei n. 12.401 de 28 de abril de 2011. Altera a Lei n. 8.080, de 19 de setembro de 1990, para dispor sobre a assistência terapêutica e a incorporação de tecnologia em saúde no âmbito do Sistema Único de Saúde – SUS. Brasília, 2011. 1 p.

BRAVATA, D. *et al.* Invariance and Inconsistency in Utility Ratings. *Medical Decision Making*, v. 25, n. 2, p. 158-167, 2005.

BRAZIER, J. *et al.* Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *British Medical Journal*, v. 305, p. 160-164, Jul. 1992.

BRAZIER, J. *et al.* A review of the use of health status measures in economic evaluation. *Health Technology Assessment*, v. 3, n. 9, p. 1-163, 1999.

BRAZIER, J.; ROBERTS, J.; DEVERILL, M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, v. 21, p. 271-292, 2002.

BRAZIER, J. *et al.* A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*, v. 13, p. 873-884, Jan. 2004.

BRAZIER, J. *et al.* Using rank and discrete choice data to estimate health state utility values on the QALY scale. *Health Economics and Decision Science, Discussion Paper Series*, n. 09/10, 2009.

BRAZIER, J. *et al.* Comparison of health states utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *Eur. J. Health Econ.*, v. 13, n. 5, p. 575-587, Oct. 2012.

BRICKMAN, P.; COATES, D.; JANOFF-BULMAN, R. "Lottery winners and accident victims: Is happiness relative?", *Journal of Personality and Social Psychology*, v. 36, p. 917-927, 1978.

CAMARGOS, M.; PERPÉTUO, I.; MACHADO, C. Expectativa de vida com incapacidade funcional em idosos em São Paulo, Brasil. *Rev. Panam. Salud Publica*. v. 17, n. 5/6, p. 379-86. 2005.

CAMERON, A.; TRIVEDI, P. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press, 2005. 1034 p.

CAMPOLINA, A.; CICONELLI, R. Qualidade de vida e medidas de utilidade: parâmetros clínicos para as tomadas de decisão em saúde. *Rev Panam Salud Publica*, v. 19, n. 2, p. 128-136, 2006.

CHEVALIER, J.; POUVOURVILLE, G. Valuing EQ-5D using Time Trade-Off in France. *The European Journal of Health Economics*, V. 14, P. 57-66, Sep. 2011.

CHUANG, L.; KIND, P. The Effect of Health State Selection on the Valuation of EQ-5D. *Medical Decision Making*, v. 31, n. 1, p. 186-194, 2010.

COAST, J. *et al.* Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics. *Journal of Health Services Research & Policy*, v. 13, n. 3, p. 31-37, Oct. 2008.

CRAIG, B.; RAMACHANDRAN, S. Relative risk of a shuffled deck: a generalizable logical consistency criterion for sample selection in health state valuation studies. *Health Economics*, v. 15, n. 8, p. 835-848, Aug. 2006.

CRAIG, B. The Duration Effect: A Link Between TTO And VAS Values. *Health Economics*, v. 18, p. 217-225, 2009.

CRAIG, B; BUSSCHABACH, J; SALOMON, J. Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. *Journal of Clinical Epidemiology*, v. 62, p. 296-305, 2009a.

CRAIG, B; BUSSCHABACH, J; SALOMON, J. Modeling Ranking, Time Trade-Off, and Visual Analog Scale Values for EQ-5D Health States: A Review and Comparison of Methods. *Medical Care*, v. 47, n. 6, p. 634-641, Jun. 2009b.

CRAIG, B. *et al.* Modeling Ranking, Time Trade-off and Visual Analogue Scale Values for EQ-5D Health States: a review and comparison of methods. *Medical Care*, v. 47, n. 6, p. 634-641, Jun. 2009.

CRUZ, L. *Medidas de Qualidade de Vida e Utilidade em uma Amostra da População de Porto Alegre*. 2010. 270 f. Tese (Doutorado em Epidemiologia) – Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

CRUZ, L. *et al.* Estimating the SF-6D Value Set for a Population-Based Sample of Brazilians. *Value in health*, v. 14, n. Suppl 1, p. S108-S114, Jul-Aug. 2011.

CUNILLERA, O. *et al.* Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Quality of Life Research*, v. 19, p. 853–864, Mar. 2010.

DEVLIN, N. *et al.* Logical inconsistencies in survey respondents' health state valuations – a methodological challenge for estimating social tariffs. *Health Economics*, v. 12, n. 7, Jul. 2003.

- DEVLIN, N.; HANSEN, P.; SELAI, C. Understanding health state valuation: A Qualitative analysis of respondents' comments. *Quality of Life Research*, v. 13, p. 1265-1277, 2004.
- DOLAN, P.; GUDEX, C. Time preference, duration and health state valuations. *Health Economics*, v. 4, n. 4, p. 289-299, Jul-Aug. 1995.
- DOLAN, P.; KIND, P. Inconsistency and Health State Valuations. *Soc. Sci. Med.*, v. 42, n. 4, p. 609-615, 1996.
- DOLAN, P. *et al.* The Time Trade-Off Method: results from a general population study. *Health Economics*, vol. 5, p. 141-154, 1996a.
- DOLAN, P. *et al.* Valuing health states: A comparison of methods. *Journal of Health Economics*, vol. 15, p. 209-231, 1996b.
- DOLAN, P. Modelling valuations for health states: the effect of duration. *Health Policy*, v. 38, p. 189-203, 1996.
- DOLAN, P. The Measurement of Health-related Quality of Life for use in Resource Allocation Decision in Health care. In: CULYER, A.; NEWHOUSE, J (Ed.). *Handbook of Health Economics*, Elsevier, 2000. v. 1B, chapter. 32, p. 1723-1760.
- DOLAN P.; ROBERTS, J. To what extent can we explain time trade-off values from other information about respondents? *Social Science & Medicine*, v. 54, p. 919-929. 2002.
- DOLAN, P. QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics*, v. 14, p. 197-208, 2005.
- DOLAN, P.; KAHNEMAN, D. Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, v. 118, p. 215-234, 2008.
- DRUMMOND, M.; STODDART, G.; TORRANCE, G. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford Medical Publications, 2<sup>nd</sup> ed., 1997.
- EFRON, B. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, v. 72, n. 359, p. 557-565, Sep. 1977.
- FEENY, D. *et al.* Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Medical Care*, v. 40, n. 2, p. 113-128, 2002.
- FERREIRA, L. Utilidades, QALYs e Medição da Qualidade de Vida. *Documento de Trabalho n. 1*. Associação Portuguesa de Economia da Saúde, Ago. 2002.
- FLYNN, T. *et al.* Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Population Health Metrics*, v. 6, Oct. 2008.
- FRANCO, G. Índices de Felicidade e Desenvolvimento Econômico. In: Ciclo de conferências "O próximo futuro: Grandes Lições", 2011, Lisboa. *Conferências...* Lisboa: Fundação Calouste Gulbenkian, 2012. p. 1-24.

GALANTE, J. *et al.* Estimation and Comparison of EQ-5D Health States' Utility Weights for Pneumococcal and Human Papillomavirus Diseases in Argentina, Chile, and the United Kingdom. *Value in Health*, v. 14, n. Suppl 1, p. S60-S64, 2011.

GOLD, M.; FRANKS, P.; ERICKSON, P. Assessing the Health of a Nation: The predictive validity of a preference-based measure and self-related health. *Medical Care*, vol. 34, n. 2, p. 163-177. Feb. 1996.

GROOT, W. Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, v. 19, n. 3, p. 403-420, 2000.

GUDEX, C. *Time Trade-Off User Manual: Props and Self-Completion Method*. Centre for Health Economics, University of York, Occasional Paper Series, 1994.

HAILEY, D.M. Health technology assessment in Canada: diversity and evolution. *The Medical Journal of Australia*, v. 187, n. 5, p. 286-88. Sep. 2007.

HAUSMAN, J.; McFADDEN, D. Specification Tests for the Multinomial Logit Model. *Econometrica*, v. 52, n. 5, p. 1219-1240, Sep. 1984.

HORNBERGER, J.; REDELMEIER, D.; PETERSEN, J. Variability Among Methods to Assess Patients' Well-Being and Consequent Effect on a Cost-Effectiveness Analysis. *Journal of Clinical Epidemiology*, v. 45, n. 5, p. 505-512, 1992.

HURLEY, J. An Overview of the normative economics of the health sector. In: CULYER, A.; NEWHOUSE, J. *Handbook of Health Economics*. Elsevier, 2000. v. 1, chapter 2, p. 56-118.

HURST, N.; RUTA, D.; KIND, P. Comparison of the Mos Short Form-12 (SF12) health status questionnaire with the SF36 in patients with rheumatoid arthritis. *British Journal of Rheumatology*, v. 37, n. 8, p. 862-869, 1998.

JACKSON, T. Health technology assessment in Australia: challenges ahead. *The Medical Journal of Australia*, v. 187, n. 5, p. 263-64. Sep. 2007.

KAHNEMAN, D.; TVERSKY, A. Prospect theory: an analysis of decision under risk. *Econometrica*, v. 47, p. 263-291, 1979.

KIND, P. *et al.* Variations in population health status: results from a United Kingdom national questionnaire survey. *BMC Health Services Research*, v. 316, n. 7133, p. 736-41, Mar. 1998.

KIND, P. *Revised Protocol for the Valuation of Health States Defined by the EQ-5D-3L Classification System: Learning the Lessons from the MVH Study*. York: Centre for Health Economics, University of York, Nov. 2009.

KLARMAN, H.; FRANCIS, J.; ROSENTHAL, G. Cost-effective analysis applied to the treatment of chronic renal disease. *Medical Care*, v. 6, n. 1, p. 46-54. Jan-Feb. 1968.

KRABBE, P.; ESSINK-BOT, M.; BONSEL, G. The Comparability and Reliability of Five Health-State Valuation Methods. *Soc. Sci. Med.*, v. 45, n. 11, p. 1641-1652, 1997.

KRABBE, P. *et al.* Testing the interval-level measurement property of multi-item visual analogue scales. *Quality of Life Research*, v. 15, n. 10, p. 1651-1661, Dec. 2006.

LAMERS, L. *et al.* Inconsistencies in TTO and VAS Values for EQ-5D Health States. *Medical Decision Making*, v. 26, n. 2, p. 173-181, Mar-Apr. 2006.

LAMERS, L. The Transformation of Utilities for Health States Worse Than Death: Consequences for the Estimation for EQ-5D Value Sets. *Medical Care*, v. 45, n. 3, p. 238-244, Mar. 2007.

LEE, Y. *et al.* South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. *Value in Health*, v. 12, n. 8, p. 1187-1193, 2009.

LEE, H. *et al.* Estimating weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. *Journal of Formosan Medical Association*, v. 112, n. 11, p. 699-706, Nov. 2013.

LOOMES, G.; McKENZIE, L. The Use of QALYs in Health Care Decision Making. *Soc. Sci. Med.*, v. 28, n. 4, p. 299-308, 1989.

LUNDENBERG, L. *et al.* Health-states utilities in a general population in relation to age, gender and socioeconomic factors. *European Journal of Public Health*, v. 9, n. 3, p. 211-217, Mar. 1999.

MACRAN, S.; WEATHRLY, H.; KIND, P. Measuring Population Health: A Comparison of Three Generic Health Status Measures. *Medical Care*, v. 41, n. 2, p. 218-231, 2003.

McCABE, C.; STEVENS, K. Visual Analogue Scales: do they have a role in the measurement of preferences for Health States? *Discussion Paper Series*, n. 04/6, Sheffield Health Economics Group, University of Sheffield, Jul. 2004.

McCABE, C. *et al.* Using ranking data to estimate health state utility models. *Journal of Health Economics*, v. 25, n. 3, p. 418-431, Feb. 2006.

McDOWELL, I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 3<sup>rd</sup> ed. New York: Oxford University Press, 2006.

McFADDEN, D. Conditional logit analysis of qualitative choice behavior. In: ZAREMBKA, P. *Frontiers in econometrics*. New York, Academic Press, 1974. p. 105-142.

MEHREZ, A.; GAFNI, A. Evaluating Health Related Quality of Life: an indifference curve interpretation for the Time Trade-off technique. *Society Science Medical*, v. 31, n. 11, p. 1281-1283, 1990.

MESQUITA, K. *Heterogeneidade Individual e o Processo de Valorização dos Estados de Saúde em Minas Gerais*. Belo Horizonte: Centro de Desenvolvimento e Planejamento Regional, UFMG, 2013. Mimeografado.

MOTA, D.; FERNANDES, M.; COELHO, H. Farmacoeconomia: um Instrumento de Eficiência para a Política de Medicamentos do Brasil. *Acta Farm. Bonaerense*, v. 22, n. 2, p. 177-186. 2003.

MURRAY, C. *et al.* Development of who guidelines on generalized cost-effectiveness analysis. *Health Economics*, v. 9, p. 235-251, 2000.

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE - NICE. Guide to the methods of technology appraisal. 2008 Jun. Available from: <<http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf>>.

NEUMANN, P. *Using cost-effectiveness analysis to improve health care: opportunities and barriers*. Oxford University Press US. 2005.

NORD, E. Methods for Quality Adjustment of Life Years. *Soc. Sci. Med.* v. 34, n. 5, p. 559-569, 1992.

PALLONI, A. *et al.* The influence of early conditions on health status among elderly Puerto Ricans. *Social Biology*, v. 52, n. 3-4, p. 132-163, Fall-Winter. 2005.

PALLONI, A. *et al.* The Tide to Come: Elderly Health in Latin America and the Caribbean. *Journal of Aging and Health*, v. 18, n. 2, p. 180-206, Apr. 2006.

PHILLIPS, C. *What is a QALY?*. Health Economics. What is...? Series, 2<sup>nd</sup> ed. UK: Hayward, 2009.

PRODUCTIVITY COMMISSION. *Impacts of Advances in Medical Technology in Australia*. Productivity Commission Research Report. Melbourne. 700p, Aug. 2005.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO - PNUD. Atlas do Desenvolvimento Humano no Brasil 2003. Disponível em: <<http://www.pnud.org.br/atlas>>.

RAND-HENDRIKSEN, K. *et al.* Time Trade-Off and Ranking Exercises are Sensitive to Different Dimensions of EQ-5D Health States. *Value in Health*, v. 15, n. 5, p. 777-782, Jul-Aug. 2012.

RATCLIFFE, J. *et al.* Using DCE and Ranking Data to Estimate Cardinal Values for Health States for Deriving a Preference-Based Single Index from the Sexual Quality of Life Questionnaire. *Health Economics*, v. 18, n. 9, p. 1261-1276, Nov. 2009.

READ, J. *et al.* Preferences for Health Outcomes: Comparison of Assessment Methods. *Medical Decision Making*, v. 4, n. 3, p. 315-329, 1984.

ROBBERSTAD, B. QALYs vs DALYs vs LYs gained: What are the differences, and what difference do they make for health care priority setting? *Norsk Epidemiologi*, v. 15, n. 2, p. 183-191, 2005.

ROBINSON, A.; DOLAN, P.; WILLIAMS, A. Valuing Health Status Using VAS and TTO: What Lies Behind the Numbers?. *Soc. Sci. Med.*, v. 45, n. 8, p. 1289-1297, 1997.

ROBINSON, A.; LOOMES, G.; JONES-LEE, M. Visual Analog Scales, Standard Gambles, and Relative Risk Aversion. *Medical Decision Making*, v. 21, n. 1, p. 17-27, Jan-Feb. 2001.

ROMERO, D.; LEITE, I.; SZWARCOWALD, C. Healthy life expectancy in Brazil: applying the Sullivan method. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 21, Sup:S7-S18, 2005.

RYAN, M.; FARRAR, S. Using conjoint analysis to elicit preferences for health care. *British Medical Journal*, v. 320, n. 3, p. 1530-1533, Jun. 2000.

RYAN, M. *et al.* Using discrete choice experiments to estimate a preference-based measure of outcome – An application to social care for older people. *Journal of Health Economics*, v. 25, n. 5, p. 927-944, Sep. 2006.

SACKETT, D.; TORRANCE, G. The utility of different health states as perceived by the general public. *Journal of Chronic Disease*, v. 31, n. 11, p. 697-704, 1978.

SALOMON, J. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Population Health Metrics*, v. 1, p. 1-12, Dec. 2003.

SCHRAMM, J. *et al.* Transição epidemiológica e o estudo de carga de doença no Brasil. *Ciência & Saúde Coletiva*, v. 9, n. 4, p. 897-908, 2004.

SECOLI, S. *et al.* Farmacoeconomia: perspectiva emergente no processo de tomada de decisão. *Ciências e Saúde Coletiva*, v. 10, n. Sup, p. 287-296. Set-Dez. 2005.

SIEGEL, S.; CASTELLAN Jr., N. *Estatística Não-Paramétrica Para Ciências do Comportamento*. 2 ed. Porto Alegre: Artmed, 2006. 448 p.

STIGGELBOUT, A. *et al.* The 'Utility' of the Visual Analog Scale in Medical Decision Making and Technology Assessment: Is It an Alternative to the Time Trade-Off?. *International Journal of Technology Assessment in Health Care*, v. 12, n. 2, p. 291-298, 1996.

SUEN, R. *Technological Advance and the Growth in Health Care Spending*. Economie D'Avant Garde Research Report No. 13, 49p. 2005.

THE WORLD BANK GROUP. World Data Bank. Education Statistics, 2013. Disponível em: <<http://databank.worldbank.org/data/>>. Acesso em: 23/05/2013.

THURSTONE, L. A Law of comparative judgment. *Psychological Review*, v. 34, p. 273-286, 1927.

TORRANCE, G.; THOMAS, W.; SACKETT, D. A utility maximization model for evaluation of health care programs. *Health Services Research*, v. 7, n. 2, p. 118-33, 1972.

TORRANCE, G. Social Preferences for Health States: an Empirical Evaluation of Three Measurement Techniques. *Socio-Econ. Plan. Sci.*, v. 10, p. 129-136, 1976.

TORRANCE, G. Measurement of Health State Utilities for Economic Appraisal: a Review. *Journal of Health Economics*, v. 5, p. 1-30, 1986.

TORRANCE, G.; FEENY, D.; FURLONG, W. Visual Analog Scales: Do They Have a Role in the Measurement of Preferences for Health States? *Medical decision Making*, v. 21, n. 4, p. 329-334, Jul-Aug. 2001.

TSUCHIYA, A. *et al.* Estimating an EQ-5D population value set: the case of Japan. *Health Economics*, v. 11, n. 4, p. 341–353, Jan. 2002.

UNITED NATIONS. Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2006 Revision. New York, 2007. Disponível em: <[http://www.un.org/esa/population/publications/wpp2006/WPP2006\\_Highlights\\_rev.pdf](http://www.un.org/esa/population/publications/wpp2006/WPP2006_Highlights_rev.pdf)>. Acesso: 23/05/2013.

van BAAL, P. *et al.* Lifetime Medical Costs of Obesity: Prevention no cure for Increasing Health Expenditure. *PLOS Medicine*, v. 5, n. 2, p. 242-249. 2008.

von NEUMANN, J.; MORGENSTERN, O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

WOOLDRIDGE, J. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press, 2002. 752 p.

WORLD HEALTH ORGANIZATION. World Health Statistics 2012. Disponível em: <[http://www.who.int/gho/publications/world\\_health\\_statistics/EN\\_WHS2012\\_Full.pdf](http://www.who.int/gho/publications/world_health_statistics/EN_WHS2012_Full.pdf)> Acesso em: 23/05/2013.

YUSOF, F.; GOH, A.; AZMI, S. Estimating an EQ-5D Value Set for Malaysia Using Time Trade-Off and Visual Analogue Methods. *Value in Health*, v. 15, n. Suppl, p. S85-S90, Jan-Feb. 2012.

ZARATE, V. *et al.* Social valuation of EQ-5D health states: the Chilean case. *Value In Health*, v. 14, p. 1135–41, 2011.

**APÊNDICE A – VALORES ESPERADOS PARA OS ESTADOS DE SAÚDE  
DEFINIDOS PELO EQ-5D**

**TABELA A - Valores esperados para os estados de saúde definidos pelo EQ-5D**

Cartão	TTO	TTO	VAS	Rank A	Rank B	Rank C	Cartão	TTO	TTO	VAS	Rank A	Rank B	Rank C
	Observado	Predito						Observado	Predito				
11112	0,840	0,884	0,846	0,928	0,911	0,945	22232	0,331	0,339	0,381	0,381	0,236	0,527
11121	0,869	0,879	0,834	0,923	0,905	0,941	22233	0,286	0,288	0,430	0,336	0,180	0,492
11122	0,783	0,817	0,774	0,851	0,816	0,886	22313	0,455	0,378	0,394	0,404	0,264	0,544
11123	0,758	0,766	0,890	0,805	0,760	0,851	22323	0,333	0,311	0,326	0,327	0,169	0,485
11211	0,819	0,850	0,792	0,911	0,890	0,932	22332	0,149	0,230	0,301	0,316	0,156	0,477
11212	0,799	0,789	0,774	0,839	0,801	0,877	22333	0,199	0,178	0,337	0,271	0,099	0,442
11221	0,795	0,783	0,764	0,834	0,794	0,873	23113	0,483	0,458	0,478	0,494	0,375	0,613
11222	0,715	0,722	0,705	0,762	0,706	0,818	23131	0,372	0,371	0,434	0,478	0,355	0,600
11223	0,640	0,670	0,608	0,716	0,650	0,783	23132	0,335	0,309	0,363	0,406	0,266	0,546
11232	0,556	0,589	0,624	0,706	0,637	0,775	23222	0,434	0,347	0,383	0,373	0,225	0,520
11312	0,665	0,679	0,648	0,774	0,721	0,827	23223	0,254	0,296	0,402	0,327	0,169	0,485
11313	0,636	0,628	0,636	0,729	0,665	0,792	23231	0,222	0,276	0,395	0,389	0,245	0,532
11323	0,602	0,561	0,575	0,651	0,569	0,733	23232	0,207	0,214	0,331	0,317	0,156	0,477
11332	0,504	0,479	0,562	0,641	0,557	0,725	23233	0,147	0,162	0,441	0,271	0,100	0,442
12111	0,794	0,824	0,775	0,878	0,849	0,907	23311	0,349	0,366	0,423	0,457	0,329	0,584
12112	0,746	0,763	0,705	0,806	0,761	0,852	23313	0,188	0,253	0,293	0,340	0,184	0,495
12121	0,755	0,757	0,691	0,801	0,754	0,847	23321	0,340	0,299	0,409	0,380	0,234	0,525
12122	0,724	0,696	0,608	0,729	0,665	0,793	23322	0,184	0,238	0,454	0,308	0,145	0,470
12123	0,655	0,644	0,584	0,683	0,609	0,758	23323	0,146	0,186	0,376	0,262	0,089	0,436
12211	0,737	0,729	0,708	0,789	0,739	0,838	23332	0,115	0,104	0,316	0,252	0,076	0,428
12212	0,688	0,667	0,616	0,717	0,650	0,783	23333	0,042	0,053	0,334	0,206	0,020	0,393
12221	0,718	0,662	0,604	0,711	0,644	0,779	31131	0,283	0,342	0,432	0,458	0,331	0,585
12312	0,646	0,558	0,593	0,652	0,570	0,734	31213	0,304	0,334	0,430	0,385	0,240	0,529
12313	0,531	0,506	0,530	0,607	0,514	0,699	31222	0,289	0,318	0,470	0,353	0,201	0,505
12331	0,437	0,419	0,534	0,591	0,494	0,687	31311	0,361	0,337	0,430	0,437	0,305	0,569
13123	0,548	0,519	0,517	0,619	0,529	0,708	31313	0,168	0,224	0,572	0,320	0,160	0,479
13211	0,614	0,604	0,647	0,724	0,660	0,789	32111	0,322	0,421	0,456	0,469	0,344	0,594
13222	0,470	0,475	0,515	0,575	0,475	0,675	32123	0,185	0,241	0,374	0,275	0,104	0,445
13232	0,317	0,342	0,478	0,519	0,406	0,632	32223	0,091	0,145	0,287	0,185	-0,006	0,377
21111	0,789	0,818	0,742	0,798	0,750	0,845	32232	0,078	0,064	0,269	0,175	-0,019	0,369
21112	0,732	0,756	0,699	0,726	0,661	0,790	32233	0,060	0,012	0,244	0,129	-0,075	0,334
21121	0,722	0,751	0,736	0,720	0,654	0,786	32322	0,171	0,087	0,308	0,166	-0,030	0,362
21122	0,718	0,689	0,678	0,648	0,566	0,731	32323	-0,006	0,036	0,382	0,120	-0,086	0,327
21123	0,569	0,638	0,592	0,603	0,510	0,696	32332	-0,038	-0,046	0,263	0,110	-0,099	0,319
21133	0,676	0,505	0,591	0,547	0,441	0,653	32333	-0,086	-0,098	0,285	0,064	-0,156	0,284
21211	0,737	0,723	0,661	0,708	0,640	0,777	33121	0,270	0,229	0,405	0,327	0,169	0,485
21212	0,657	0,661	0,593	0,637	0,551	0,722	33122	0,263	0,167	0,364	0,256	0,081	0,430
21221	0,679	0,656	0,563	0,631	0,544	0,718	33211	0,223	0,200	0,390	0,315	0,155	0,476
21231	0,482	0,522	0,643	0,575	0,475	0,675	33213	0,065	0,087	0,311	0,198	0,010	0,386
21311	0,684	0,613	0,682	0,643	0,560	0,727	33221	0,092	0,133	0,284	0,238	0,059	0,417
21312	0,563	0,551	0,545	0,572	0,471	0,672	33222	0,038	0,071	0,294	0,166	-0,030	0,362
21313	0,575	0,500	0,491	0,526	0,415	0,637	33223	0,039	0,020	0,212	0,121	-0,086	0,327
21331	0,531	0,413	0,487	0,510	0,395	0,625	33231	0,031	0,000	0,302	0,182	-0,010	0,374
21332	0,402	0,351	0,484	0,438	0,306	0,570	33232	0,023	-0,062	0,248	0,110	-0,099	0,319
22111	0,693	0,697	0,653	0,676	0,599	0,752	33233	-0,055	-0,113	0,237	0,065	-0,155	0,284
22112	0,615	0,635	0,554	0,604	0,511	0,697	33312	0,108	0,029	0,291	0,179	-0,014	0,372
22113	0,583	0,583	0,530	0,558	0,454	0,662	33313	0,048	-0,023	0,295	0,133	-0,071	0,337
22121	0,617	0,630	0,560	0,598	0,504	0,693	33322	-0,070	-0,038	0,223	0,102	-0,110	0,312
22211	0,628	0,601	0,526	0,586	0,489	0,683	33323	-0,046	-0,090	0,202	0,056	-0,166	0,278
22221	0,510	0,534	0,457	0,509	0,394	0,624	33333	-0,235	-0,223	0,137	0,000	-0,235	0,235

Fonte: elaboração do autor.

**APÊNDICE B – EXTRATO DA MATRIZ DE DOMINÂNCIA LÓGICA**

**TABELA B – Extrato da Matriz de Dominância Lógica**

	11111	11112	11121	11122	11123	11211	11212	11221	33233	33312	33313	33322	33323	33333
11111	0	1	1	1	1	1	1	1	1	1	1	1	1	1
11112	-1	0	0	1	1	0	1	0	1	1	1	1	1	1
11121	-1	0	0	1	1	0	0	1	1	0	0	1	1	1
11122	-1	-1	-1	0	1	0	0	0	1	0	0	1	1	1
11123	-1	-1	-1	-1	0	0	0	0	1	0	0	0	1	1
11211	-1	0	0	0	0	0	1	1	1	1	1	1	1	1
11212	-1	-1	0	0	0	-1	0	0	1	1	1	1	1	1
11221	-1	0	-1	0	0	-1	0	0	1	0	0	1	1	1
11222	-1	-1	-1	-1	0	-1	-1	-1	1	0	0	1	1	1
11223	-1	-1	-1	-1	-1	-1	-1	-1	1	0	0	0	1	1
11232	-1	-1	-1	-1	0	-1	-1	-1	1	0	0	0	0	1
11312	-1	-1	0	0	0	-1	-1	0	0	1	1	1	1	1
11313	-1	-1	0	0	0	-1	-1	0	0	0	1	0	1	1
11323	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1
33122	-1	-1	-1	-1	0	0	0	0	1	0	0	1	1	1
33211	-1	0	0	0	0	-1	0	0	1	1	1	1	1	1
33213	-1	-1	0	0	0	-1	-1	0	1	0	1	0	1	1
33221	-1	0	-1	0	0	-1	0	-1	1	0	0	1	1	1
33222	-1	-1	-1	-1	0	-1	-1	-1	1	0	0	1	1	1
33223	-1	-1	-1	-1	-1	-1	-1	-1	1	0	0	0	1	1
33231	-1	0	-1	0	0	-1	0	-1	1	0	0	0	0	1
33232	-1	-1	-1	-1	0	-1	-1	-1	1	0	0	0	0	1
33233	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	1
33312	-1	-1	0	0	0	-1	-1	0	0	0	1	1	1	1
33313	-1	-1	0	0	0	-1	-1	0	0	-1	0	0	1	1
33322	-1	-1	-1	-1	0	-1	-1	-1	0	-1	0	0	1	1
33323	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	1
33333	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0

Fonte: elaboração do autor.

## APÊNDICE C – COMPARAÇÃO ENTRE TTO E RANK C

**TABELA C - Resultados do modelo logit condicional (Ordenação) e do modelo de efeitos aleatórios (Troca de Tempo)**

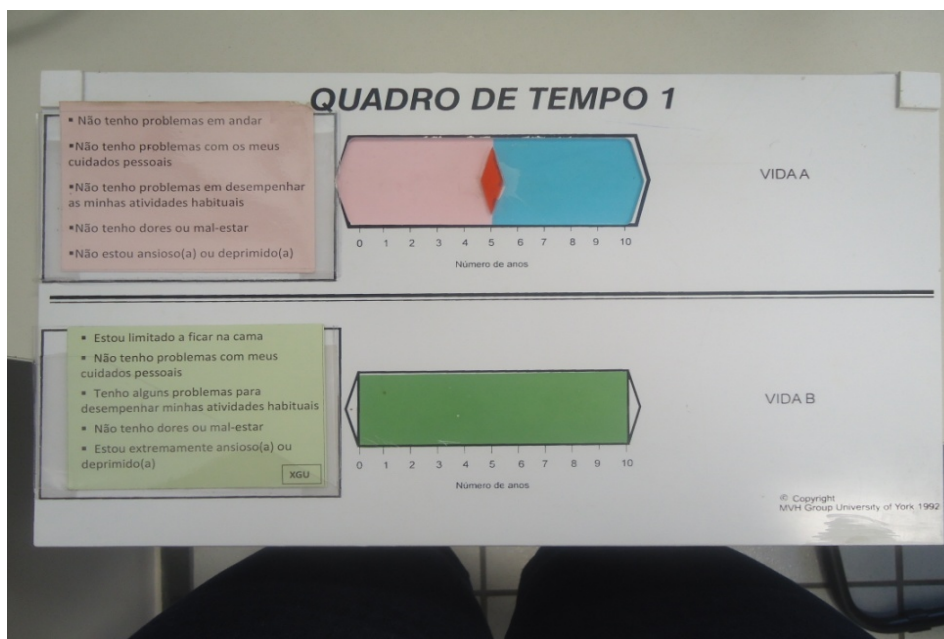
<b>Variáveis</b>	<b>Coefficientes TTO</b>	<b>dp</b>	<b>Coefficientes Rank</b>	<b>dp</b>
Mobilidade 2	0,127***	0,007	1,140***	0,029
Cuidados Pessoais 2	0,121***	0,007	0,686***	0,026
Atividades Habituais 2	0,095***	0,007	0,512***	0,026
Dor/Desconforto 2	0,067***	0,007	0,435***	0,022
Ansiedade/Depressão 2	0,061***	0,007	0,406***	0,025
Mobilidade 3	0,403***	0,008	2,261***	0,045
Cuidados Pessoais 3	0,246***	0,008	1,029***	0,031
Atividades Habituais 3	0,205***	0,008	0,859***	0,031
Dor/Desconforto 3	0,200***	0,007	0,729***	0,026
Ansiedade/Depressão 3	0,113***	0,007	0,651***	0,027
Morte			7,373***	0,091
n. estados		102		102
Diferença Absoluta Média		0,035		0,189
No. >0.05		25		95
No. >0.10		2		73
rho de Pearson		0,987		0,973
rho de Spearman		0,987		0,972

Fonte: Elaboração do autor.

\*\*\* Significante a 1%.

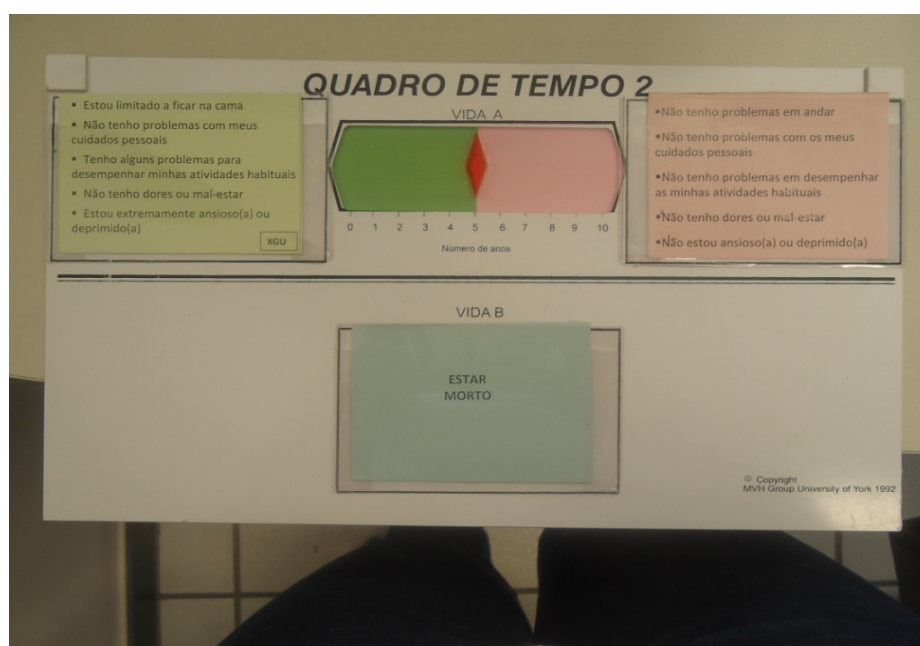
## APÊNDICE D – QUADRO DA TROCA DE TEMPO

**FIGURA A – QUADRO DE TEMPO 1: Estados melhores que a morte**

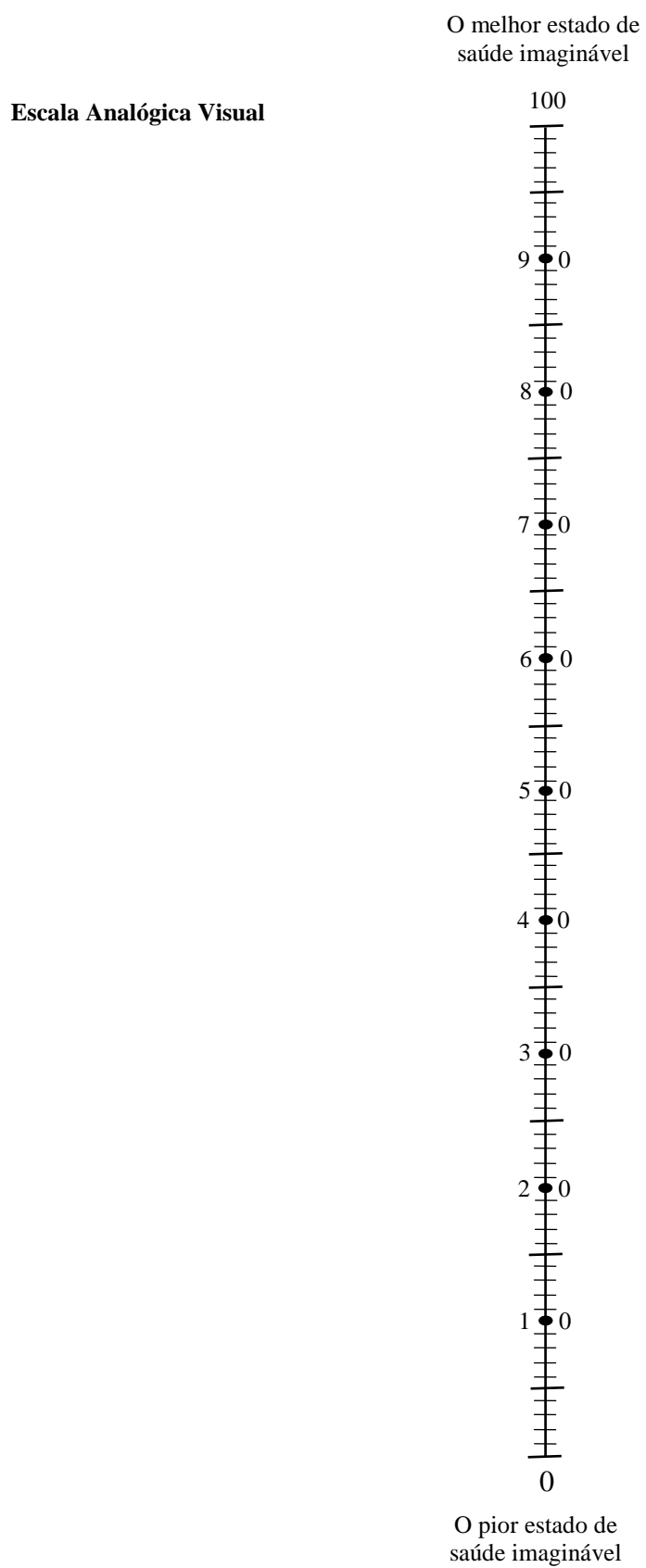


Fonte: Mesquita, 2013.

**FIGURA B – QUADRO DE TEMPO 2: Estados piores que a morte**



Fonte: Mesquita, 2013.

**ANEXO A - INSTRUMENTOS**

**SISTEMA DESCRITIVO EQ-5D**

Assinale com uma cruz (assim ☒), um quadrado de cada um dos seguintes grupos, indicando qual das afirmações melhor descreve o seu estado de saúde **atual**.

**Mobilidade**

- Não tenho problemas em andar
- Tenho alguns problemas em andar
- Estou limitado a ficar na cama

**Cuidados Pessoais**

- Não tenho problemas com os meus cuidados pessoais
- Tenho alguns problemas para me lavar ou me vestir
- Sou incapaz de me lavar ou vestir sozinho

**Atividades Habituais** (*ex. trabalho, estudos, atividades, domésticas, atividades em família ou de lazer*)

- Não tenho problemas em desempenhar as minhas atividades habituais
- Tenho alguns problemas em desempenhar as minhas atividades habituais
- Sou incapaz de desempenhar as minhas atividades habituais

**Dor/Mal Estar**

- Não tenho dores ou mal-estar
- Tenho dores ou mal-estar moderados
- Tenho dores ou mal-estar extremos

**Ansiedade/Depressão**

- Não estou ansioso(a) ou deprimido(a)
- Estou moderadamente ansioso(a) ou deprimido(a)
- Estou extremamente ansioso(a) ou deprimido(a)

## SISTEMA DESCRITIVO EQ-5D

- Eu vou fazer algumas perguntas para você sobre diferentes estados de saúde e doença.
- Não existem respostas certas ou erradas, eu apenas gostaria de saber o que you pensa.
- Tudo o que você disser nesta entrevista será mantido em sigilo.

**ENTREGUE O SISTEMA DESCRITIVO EQ-5D AO ENTREVISTADO E PEÇA AO ENTREVISTADO QUE O PREENCHA**

### 1.

- Primeiramente, eu vou lhe pedir para que pense sobre seu estado de saúde atual.
- Marque um X em cada um dos grupos para mostrar quais situações melhor descrevem sua saúde atual.
- Assinale com uma cruz (assim ) , um quadrado de cada um dos seguintes grupos, indicando qual das afirmações melhor descreve o seu estado de saúde **atual**.

**PEGUE O ENVELOPE DO GRUPO DE CARTÕES E EMBARALHE-OS.**

## ORDENAÇÃO

### 2.

- Aqui estão alguns cartões, cada um descreve um estado de saúde.
- Cada estado de saúde é descrito por uma combinação de cinco características deste questionário que você utilizou para avaliar o seu estado de saúde e o cartão referente a “ESTAR MORTO”.
- Os cartões não estão ordenados.
- Eu quero que você os ordene, deixando na parte superior o estado de saúde que você avalia ser o melhor, e na parte inferior o estado de saúde que você avalia como sendo o pior.
- Se você acha que alguns destes estados são iguais, posicione-os um ao lado do outro.
- Para cada um desses estados, eu quero que você se imagine nestas situações.
- Para cada estado, pense que a situação descrita vai durar por 10 anos, sem qualquer mudança, e depois você morre.
- Eu gostaria que você os ordenasse juntamente com o cartão em que está escrito “ESTAR MORTO”.
- Se você achar que algum estado é pior do que estar morto, posicione este cartão abaixo do cartão “ESTAR MORTO”.
- Agora, por favor, ordene os cartões.


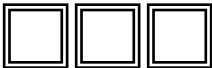
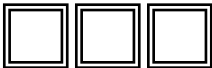
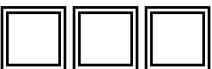
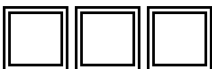
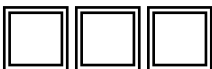
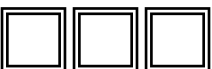
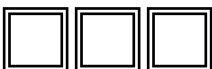
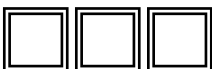

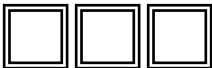
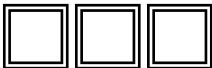
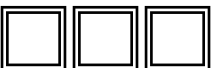
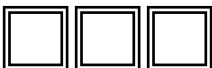
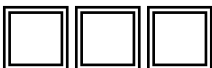
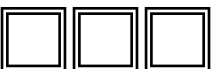
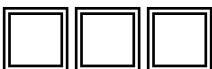
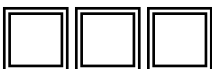
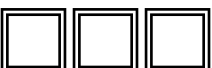
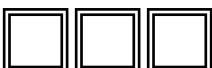
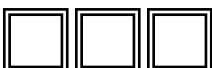
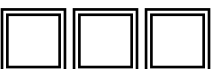
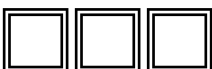
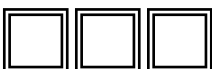
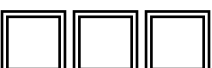
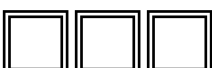
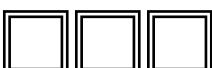
**QUANDO O ENTREVISTADO TERMINAR DE ORDENAR OS CARTÕES, VERIFIQUE SE ELE ENCONTRA-SE SATISFEITO COM A ORDENAÇÃO.**

- Você está satisfeito com a sua ordenação? Se não estiver, fique à vontade para alterá-la.

**DEIXE OS CARTÕES SOBRE A MESA EM SUA ORDEM DE CLASSIFICAÇÃO. REGISTRE A ORDEM DOS ESTADOS NA COLUNA COM CABEÇALHO “MELHOR SITUAÇÃO”, NO QUADRO DE CLASSIFICAÇÃO I.**

A) SE SOMENTE UM CARTÃO É ALOCADO PARA DETERMINADA POSIÇÃO, ESCREVA O CÓDIGO DE LETRAS CORRESPONDENTE NO ESPAÇO TRIPLO.

B) SE DOIS OU MAIS CARTÕES SÃO ALOCADOS NA MESMA POSIÇÃO, ESCREVA OS CÓDIGOS DE LETRAS DESSES CARTÕES NA MESMA LINHA.

<b>QUADRO DE CLASSIFICAÇÃO I</b>		
MELHOR SITUAÇÃO	MELHOR SITUAÇÃO	MELHOR SITUAÇÃO
1. 		
2. 		
3. 		
4. 		
5. 		
6. 		
7. 		
8. 		
9. 		
PIOR SITUAÇÃO	PIOR SITUAÇÃO	PIOR SITUAÇÃO

**RECOLHA OS CARTÕES DA MESA.**

**ESCALA ANALÓGICA VISUAL (EAV)**

**ENTREGUE A ESCALA ANALÓGICA VISUAL AO ENTREVISTADO**

3.

- Eu gostaria que você olhasse esta escala.

- Como você pode ver, a extremidade superior marca 100.
- Este é o melhor estado de saúde possível que você pode imaginar.
- A extremidade inferior marca 0 e este é o pior estado de saúde possível que você pode imaginar.
- Agora eu gostaria que você marcasse na escala o quão bom ou ruim você pensa ser seu estado de saúde atual.
- Escreva “EU” próximo a este traço.

**REGISTRE A PONTUAÇÃO QUE O ENTREVISTADO MARCOU PARA AVALIAR SEU PRÓPRIO ESTADO DE SAÚDE.**



**RETIRE O CARTÃO “ESTAR MORTO”, EMBARALHE OS DEMAIS CARTÕES E ENTREGUE-OS AO ENTREVISTADO**

**4.**

- Agora você vai marcar onde estes estados deveriam estar na escala.
- Imagine ainda que você se encontra em cada um desses estados e que ele vai durar por 10 anos, sem qualquer mudança, e então você morre.
- Use a escala de modo que os estados que você considera parecidos tenham pontuações próximas e os estados que você considera muito diferentes entre si tenham pontuações mais distantes.
- Você pode dar a mesma pontuação para mais de um estado.
- Escreva o código de letras de cada cartão próximo a cada traço que você realizar.

**DEPOIS QUE O ENTREVISTADO TERMINAR DE VALORAR OS CARTÕES ENTREGUE O CARTÃO “ESTAR MORTO”.**

**5.** Aqui está o cartão “ESTAR MORTO”. Onde você o colocaria nesta escala? Você pode alterar o local na escala em que você colocou os outros estados, caso queira.

**RECOLHA A ESCALA ANALÓGICA VISUAL**

**REGISTRE A PONTUAÇÃO DE CADA CARTÃO AO LADO DOS SEUS RESPECTIVOS CÓDIGOS DE LETRAS NA COLUNA EAV DO QUADRO DE CLASSIFICAÇÃO II.**

**QUADRO DE CLASSIFICAÇÃO II**

REGISTRO DO CÓDIGO DO CARTÃO E PONTUAÇÃO NO ESPAÇO EAV.

**CÓDIGO DO CARTÃO****EAV**1.  2.  3.  4.  5.  6.  7.  8.  9.

## PROTOCOLO DA TROCA DE TEMPO

**RETIRE OS CARTOES DE SAÚDE PERFEITA E ESTAR MORTO E EMBARALHE OS DEMAIS.**

**MOSTRE O LADO 1 DO QUADRO DA TROCA DE TEMPO**

- Eu vou mostrar para você algumas descrições dos estados de saúde que você viu anteriormente.
- Eu quero que você imagine que VOCÊ está nesses estados, os quais podem permanecer por diferentes durações.
- Não há respostas certas ou erradas. Eu só vou querer saber qual VOCÊ escolhe.

**MOSTRE O CARTÃO COM O ESTADO PARA O RESPONDENTE.**

- Por favor, leia a descrição deste cartão.

**REGISTRE O CÓDIGO DO CARTÃO REFERENTE A ESTE ESTADO DE SAÚDE NOS QUADRADOS ABAIXO**

1)

**USE O QUADRO DA TROCA DE TEMPO “1”  
INSIRA O CARTÃO NO BOLSO CORRESPONDENTE À “VIDA B” E MOVA O MARCADOR DO QUADRO DA TROCA DE TEMPO DA “VIDA A” PARA “0” ANO**

- Vida B significa viver 10 anos neste estado e depois você morre
- Vida A não dá a você nenhum tempo de vida – você morreria agora
- Qual das duas vidas você prefere? Vida A ou Vida B ou elas são a mesma coisa?

**REGISTRE A RESPOSTA DO INDIVÍDUO COM UM X NOS CAMPOS AO LADO DO QUADRO ABAIXO**

2)

Escolha	Significado	Ação	SEÇÃO	
VIDA A	Estado Pior que a morte	Vá para Pior que a Morte	II	<input type="checkbox"/>
VIDA B	Estado melhor que a morte	Vá para Melhor que a Morte	I	<input type="checkbox"/>
Mesma Coisa	Ponto de Indiferença Estabelecido	FIM		<input type="checkbox"/>

**SE O INDIVÍDUO ESCOLHEU VIDA “A” VÁ PARA A SEÇÃO II DO QUESTIONÁRIO  
SE O INDIVÍDUO ESCOLHEU VIDA “B” VÁ PARA A SEÇÃO I DO QUESTIONÁRIO  
SE VIDA “A” E VIDA “B” REPRESENTAM A MESMA COISA PARA O INDIVÍDUO, ENCERRE A AVALIAÇÃO DESTE ESTADO E PASSE PARA O PRÓXIMO CARTÃO.**

### QUADRO DE VALOR

VALOR (anos)	<input type="text"/>
--------------	----------------------

## SEÇÃO I – ESTADO MELHOR QUE A MORTE

**CONTINUE USANDO O QUADRO DE TROCA DE TEMPO 1**

**REGULE O MARCADOR DO QUADRO DA TROCA DE TEMPO PARA 5 ANOS ( $t = 5$ )**

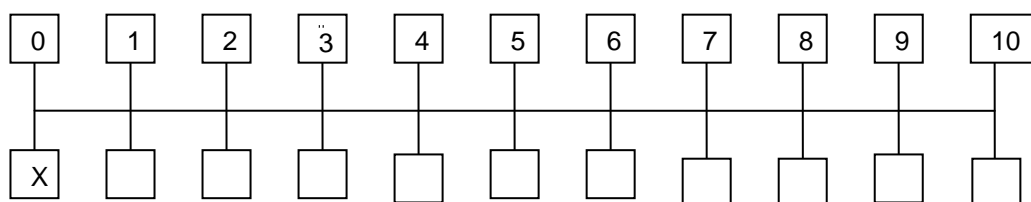
- Vida A significa que você vive por  $t$  anos em saúde perfeita e depois morre.
- Vida B significa que você vive por 10 anos nesse estado de saúde e depois morre.
- Você preferiria a Vida A ou Vida B ou elas são a mesma coisa?

Escolha	Significado	Ação
VIDA A	Prefere $t$ anos em saúde perfeita a 10 anos em pior estado de saúde	Marque um ✓ embaixo de $t$ Mova o marcador do quadro da troca de tempo para <b>DIMINUIR</b> 1 ano ( $\downarrow t$ )
VIDA B	Prefere 10 anos em pobre estado de saúde a $t$ anos em saúde completa	Marque um X embaixo de $t$ Mova o marcador do quadro da troca de tempo para <b>AUMENTAR</b> 1 ano ( $\uparrow t$ )
A mesma coisa	Ponto de indiferença estabelecido	Marque um sinal de igualdade (=) embaixo de $t$

**SE UM ✓ E UM X OCORREREM UM AO LADO DO OUTRO ENCERRE A AVALIAÇÃO DESTE ESTADO DE SAÚDE E REGISTRE O VALOR DO PONTO MÉDIO NO QUADRO DE VALOR.**

**SE UM = FOR REGISTRADO ENCERRE A AVALIAÇÃO DESTE ESTADO DE SAÚDE E REGISTRE O VALOR DO  $t$  NO QUADRO DE VALOR.**

**CASO CONTRÁRIO REPITA COM O NOVO TEMPO  $t$ .**



**QUADRO DE VALOR**

VALOR (anos)	
--------------	--

**INICIE A AVALIAÇÃO DO PRÓXIMO ESTADO DE SAÚDE ATÉ COMPLETAR OS 7 ESTADOS DE SAÚDE**

## SEÇÃO II – ESTADO PIOR QUE A MORTE

**PASSE A UTILIZAR O QUADRO DA TROCA DE TEMPO 2**

**RETIRE O CARTÃO DO ESTADO AVALIADO DO LADO 1 E PASSE-O PARA O LADO 2**

**REGULE O MARCADOR DO QUADRO DA TROCA DE TEMPO PARA 5 ANOS ( $t = 5$ )**

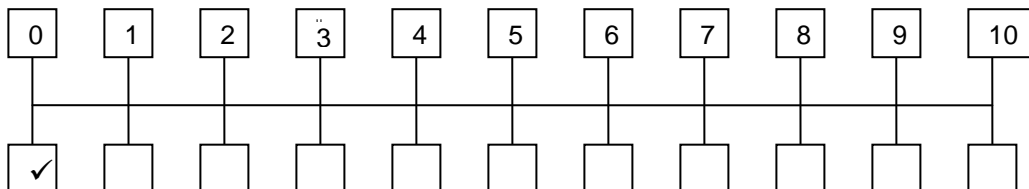
- Vida A significa que você viveria por  $t$  anos neste estado de saúde, seguido de  $10-t$  anos em saúde perfeita.
- Vida B significa estar morto.
- Você preferiria a Vida A ou a Vida B ou elas são a mesma coisa?

Escolha	Significado	Ação
VIDA A	Prefere $t$ anos em saúde ruim e acrescido de $10-t$ em saúde perfeita a estar morto	Marque um <input checked="" type="checkbox"/> embaixo de $t$ Mova o marcador do quadro da troca de tempo para <b>AUMENTAR</b> 1 ano ( $\uparrow t$ )
VIDA B	Prefere estar morto a viver em estado de saúde ruim	Marque um <input type="checkbox"/> embaixo de $t$ Mova o marcador do quadro da troca de tempo para <b>DIMINUIR</b> 1 ano ( $\downarrow t$ )
A mesma coisa	Ponto de indiferença estabelecido	Marque um sinal de igual (=) embaixo de $t$

**SE UM  E UM  OCORREREM UM AO LADO DO OUTRO ENCERRE A AVALIAÇÃO DESTE ESTADO DE SAÚDE E REGISTRE O VALOR DO PONTO MÉDIO NO QUADRO DE VALOR.**

**SE UM = FOR REGISTRADO ENCERRE A AVALIAÇÃO DESTE ESTADO DE SAÚDE E REGISTRE O VALOR DO  $t$  NO QUADRO DE VALOR.**

**CASO CONTRÁRIO REPITA COM O NOVO TEMPO  $t$ .**



**QUADRO DE VALOR**

VALOR (anos)	
--------------	--

**INICIE A AVALIAÇÃO DO PRÓXIMO ESTADO DE SAÚDE ATÉ COMPLETAR OS 7 ESTADOS DE SAÚDE**