

COMPARAÇÃO DE MÉTODOS DE VALORAÇÃO DE PREFERÊNCIAS SOCIAIS POR ESTADOS DE SAÚDE

Carla de Barros Reis¹

Mônica Viegas Andrade²

Kenya Noronha³

Este artigo explora a equivalência entre três métodos de aferição de preferências por estados de saúde: Ranking, Visual Analogue Scale (VAS) e Time Trade-off (TTO). As preferências são utilizadas em avaliações de tecnologias em saúde. Utilizando dados de um estudo de avaliação com base no sistema descritivo EQ-5D, em Minas Gerais, realiza-se a transformação das informações ordinais em cardinais por meio do modelo *logit* condicional. Os resultados mostram que apesar da ordenação das preferências ser similar entre os três métodos analisados, diferenças expressivas na magnitude dos valores são observadas, com consequências importantes para a tomada de decisões quanto à alocação de recursos em saúde.

Palavras-chave: EQ-5D; métodos de valoração; preferências; avaliação de tecnologias em saúde.

COMPARISON OF VALUATION METHODS OF HEALTH STATES SOCIETAL PREFERENCES

This paper aims to analyze the equivalence between three valuation methods for eliciting health states preferences: Ranking, Visual Analogue Scale (VAS) and Time Trade-off (TTO). Preferences are used in health technology assessment. Data come from a survey in Minas Gerais using the EQ-5D descriptive system. The transformation of ordinal preferences into cardinal preferences was performed using the conditional logit model. The results show that despite the ranking of preferences to be similar among the three methods, significant differences in values are observed and play important consequences for making decision regarding allocation of health resources.

Keywords: EQ-5D; valuation methods; health states preferences; health technology assessment.

JEL: I31; I18; D61.

1 INTRODUÇÃO

Preferências individuais por estados de saúde são requeridas quando a introdução ou substituição de uma tecnologia, tratamento ou medicamento em saúde é avaliada no contexto das análises custo-utilidade. As pressões para a contenção de dispêndios com cuidados em saúde e a necessidade de uma alocação mais eficiente dos recursos destinados ao setor têm destacado a importância dessas análises no auxílio das tomadas de decisões (Brasil, 2008; Drummond; Stoddart; Torrance, 1997). Contudo, uma fonte inicial de resistência ao uso de análises custo-utilidade refere-se à validade e à comparabilidade dos resultados entre estudos (Luce, 2005;

1. Escola Nacional de Saúde Pública Sergio Arouca/Fundação Oswaldo Cruz.

2. Centro de Desenvolvimento e Planejamento Regional/Universidade Federal de Minas Gerais.

3. Centro de Desenvolvimento e Planejamento Regional/Universidade Federal de Minas Gerais.

Drummond e Sculpher, 2005; McDonough e Tosteson, 2007). Questões, como baixa comparabilidade e fraca normalização entre avaliações, possuem potencial para enfraquecer o objetivo fundamental das análises de custo-utilidade, qual seja, balizar a tomada de decisões.

Resultados conflitantes podem refletir a variabilidade de uma série de fatores, como as diferenças de amostras analisadas, sistemas descritivos de estados de saúde, metodologias de obtenção de preferências e procedimentos de pesquisa. No que se refere aos instrumentos de avaliação de qualidade de vida, pode-se escolher entre aqueles classificados como genéricos ou específicos. Os primeiros são aplicáveis em diferentes contextos e intervenções em saúde, já que possibilitam uma avaliação simultânea de várias dimensões, podem ser utilizados em qualquer população e permitem comparações entre pacientes com diferentes patologias e morbidades. Por sua vez, os sistemas específicos são designados a condições específicas de doença (diabetes, artrite, 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).

No que se refere à decisão sobre quais preferências devem ser utilizadas na determinação dos benefícios em avaliações de custo-utilidade, a escolha ocorre entre a participação de pacientes que experimentam uma dada condição de saúde *versus* uma amostra representativa da sociedade em geral. No primeiro caso, argumenta-se que os pacientes e médicos são os indivíduos que, de fato, possuem conhecimento do estado de saúde em questão e, portanto, expressam a verdadeira perda de utilidade que este gera no bem-estar individual. A defesa pelo uso das preferências do público geral baseia-se no fato de que a sociedade é responsável pelo pagamento de impostos. Nesse sentido, a sociedade não deve se responsabilizar apenas pelo financiamento das políticas e programas de saúde, mas deve também participar da determinação e alocação dos benefícios. A literatura sugere que, devido ao processo de adaptação experimentado por indivíduos acometidos por doenças ou situações adversas de saúde, pacientes atribuem maiores utilidades a estados de saúde precários do que ao público geral (McTaggart-Cowan *et al.*, 2010; Dolan, 1999). Gabriel *et al.* (1999) buscam determinar se a escolha da amostra de estudos de obtenção de preferências possui impactos sobre a análise de custo-utilidade de intervenções para o tratamento de osteoporose. As utilidades obtidas por meio de mulheres que não sofreram fraturas osteoporóticas são aproximadamente 50% menores que aquelas geradas a partir de mulheres que experimentaram essa condição de saúde. As diferenças de magnitude poderiam alterar as estimativas de custo-utilidade.

No que tange à escolha do método de obtenção de preferências em saúde, o debate refere-se, principalmente, à priorização de técnicas baseadas na troca. Essas técnicas incorporam o conceito microeconômico de utilidade esperada

e conduzem à realização de tarefas de difícil administração, com alto nível de abstração, além da adoção de algumas hipóteses teóricas implícitas em sua construção (Lamers, 2007; Robison, Loomes e Jones-Lee, 2001; Craig, 2009). Outras métricas conseguem refletir relações de preferência com maior factibilidade e simplicidade na coleta das informações, mas não encontram base na teoria microeconômica da escolha (McCabe *et al.*, 2006; Ratcliffe *et al.*, 2009). Estudos com pacientes acometidos por doença renal crônica ilustra a grande dificuldade em encontrar um instrumento padrão para avaliar qualidade de vida relacionada à saúde. A utilização de diferentes técnicas resulta em desfechos discrepantes. A partir da comparação de seis métodos de valoração de preferências por estados de saúde, Hornberger, Redelmeier e Petersen (1992) encontram diferenças significativas entre os níveis de bem-estar de pacientes sob tratamento de hemodiálise, com consequências diretas sobre as estimativas de custo por benefício ganho. Lenert *et al.* (1998) e Dolan e Sutton (1997) corroboram o efeito da escolha de diferentes metodologias sobre os resultados. Os autores também examinam o peso da variação das formas de administração dos questionários sobre os valores médios de utilidades. Os resultados sugerem que a maneira com que as questões são estruturadas no questionário pode ter um efeito significativo sobre os resultados e que as diferenças entre as preferências obtidas utilizando distintas variantes de um mesmo método apresentam-se maiores que aquelas encontradas por distintos métodos de valoração.

Apesar de algumas abordagens terem sido sugeridas (Torrance, 1986), não existe um consenso quanto à melhor maneira de se determinar a qualidade de vida relacionada à saúde. Atualmente, o serviço de saúde público americano, por meio do *United States Panel on Cost-Effectiveness in health and Medicine*, recomenda o uso do Caso de Referência como padrão metodológico para análises de custo-utilidade. Nessa perspectiva, as avaliações devem utilizar sistemas descritivos genéricos, métodos de valoração baseados na teoria microeconômica e preferências de uma amostra representativa da população geral. Desde a publicação do Caso de Referências, em 1996, observa-se uma maior qualidade e aceitabilidade no uso das avaliações para respaldar tomada de decisões no sistema de saúde americano (Neumann *et al.*, 2005). O estabelecimento de critérios para a Avaliação de Tecnologias em Saúde (ATS) faz-se mais fortemente nos países com tradição em análises de custo-efetividade. No âmbito internacional, a política de avaliação está consolidada em países como a Austrália, o Canadá e a Inglaterra, dado que as análises de custo-efetividade foram explicitamente incorporadas no processo decisório das políticas públicas desses países (Brasil, 2008; Jackson, 2007; Neumann, 2005). Na Inglaterra, as avaliações de tecnologias em saúde foram introduzidas no processo decisório pela criação do National Institute for Health and Clinical Excellence (Nice), em 1999. O Nice possui objetivos 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 nesse 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). Assim como no caso americano, o Nice recomenda a utilização de um método baseado na escolha (TTO) para um conjunto de estados de saúde classificados de acordo com um instrumento genérico – o EQ-5D (Nice, 2008).

No Brasil, as avaliações de tecnologias em saúde vêm sendo, paulatinamente, incorporadas aos processos de tomada de decisão, à medida que a carência de recursos exige que sejam alcançados resultados mais eficientes (Brasil, 2008). 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). Avanços também têm sido conquistados com a criação da Rede Brasileira de Avaliação de Tecnologias em Saúde (Rebrats), em 2008, e da Comissão Nacional de Incorporação de Tecnologias no SUS (Conitec), em 2011. Essas instituições estabelecem critérios e prazos para a realização de análises, além de manter a responsabilização pela avaliação de inovações tecnológicas que vierem a ser utilizadas no sistema público de saúde (Andrade *et al.*, 2013). Apesar do relativo avanço experimentado nos últimos anos com respeito aos marcos institucionais ligados à ATS no Brasil, muitos desafios ainda se colocam para a garantia da qualidade dos estudos. As informações disponibilizadas pelo Sistema de Informação da Rede Brasileira de Avaliação de Tecnologias em Saúde (Sisrebrats) sinalizam algumas dessas dificuldades. Entre 2008 e 2013, dos 333 estudos registrados no Sisrebrats, pouco mais de 10% representam avaliações econômicas e, entre as análises custo-utilidade, todas fazem uso de preferências obtidas em países desenvolvidos (Brasil, 2015). Nesse sentido, aspectos mais detalhados no que se referem às práticas e às metodologias a serem adotadas nas análises nacionais ainda representam uma questão aberta à discussão.

O objetivo deste artigo é comparar as utilidades geradas por três diferentes métodos de valoração de preferências por estados de saúde – Ranking, Visual Analogue Scale (VAS) e Time Trade-Off (TTO) – utilizando uma pesquisa realizada para o estado de Minas Gerais. Além disso, pretende-se inferir como esses resultados podem afetar o processo decisório de tratamentos e programas de saúde no âmbito do Sistema Único de Saúde (SUS). Reconhecendo a existência de uma gama de fatores responsáveis pela variabilidade de resultados, mantêm-se fixas algumas recomendações observadas na literatura internacional, com a adoção de uma amostra representativa da população mineira e a adoção de um sistema descritivo genérico de estados de saúde – o EQ-5D. O único elemento a sofrer variações é a metodologia de obtenção de preferências sociais. As comparações entre os métodos são realizadas a partir dos Coeficientes de Spearman, Pearson e Kendall tau, além do cálculo das diferenças absolutas médias entre os valores esperados. Os resultados mostram que apesar da ordenação

das preferências por estados de saúde ser bastante similar entre os três métodos analisados, diferenças expressivas na magnitude dos valores são observadas. Essas diferenças cardinais podem trazer importantes consequências para a tomada de decisões quanto à alocação de recursos em saúde.

2 MÉTODOS DE OBTENÇÃO DE PREFERÊNCIAS

As avaliações que ponderam os custos monetários e os benefícios associados a uma tecnologia em saúde, considerando as preferências sociais, são conhecidas como análises de custo-utilidade. A métrica usual dos benefícios nesse tipo de análise é o uso dos Anos de Vida Ajustados pela Qualidade (Avaq), que combinam informações sobre mortalidade e qualidade de vida relacionada à saúde (Phillips, 2009). Para a construção do Avaq é necessário conhecer a sequência de estados de saúde associada a cada doença e a forma como a sociedade valoriza determinadas condições de saúde, ou seja, as preferências sociais por estados de saúde. As preferências têm sido aferidas por meio do uso de dois métodos fundamentados na teoria da utilidade esperada: *Time trade off* (TTO) e *Standard Gamble* (SG). Esses métodos incorporam a ideia de *trade-off* entre qualidade de vida e longevidade, tendo como pressuposto que os indivíduos são capazes de estabelecer a taxa marginal de substituição entre qualidade e quantidade.

No método *Standard Gamble*, para valorar cada estado de saúde, o indivíduo realiza uma escolha entre duas loterias. A primeira refere-se a um tratamento com dois resultados possíveis: com probabilidade, o tratamento é bem-sucedido e o paciente retorna à saúde perfeita permanecendo nesse estado por t anos adicionais; com probabilidade $1-p$, o tratamento é mal-sucedido e o paciente morre. Na segunda loteria, o indivíduo não recebe tratamento e permanece naquele estado de saúde hipotético por t anos. O exercício de escolha é realizado até que se encontre uma probabilidade p que torne o respondente indiferente entre as duas loterias (Furlong *et al.*, 1990). O método *Time Trade-off* foi desenvolvido como uma alternativa ao SG. No TTO o indivíduo realiza uma escolha entre duas alternativas que envolvem a troca entre qualidade e tempo de vida. A diferença entre os dois métodos é que no TTO a escolha é determinística enquanto no SG a escolha ocorre entre duas loterias. No TTO, o indivíduo escolhe entre duas alternativas. Na primeira, o indivíduo viveria com uma determinada condição de saúde por t anos seguidos de morte, enquanto na segunda, o respondente viveria por $x < t$ anos com saúde perfeita, seguidos por morte. Dessa maneira, a escolha do indivíduo envolve a troca direta de tempo por qualidade de vida. O exercício é realizado até que se encontre um número de anos em saúde perfeita (x) que torne o respondente indiferente entre as duas alternativas. Por incorporarem a ideia de sacrifício ou troca entre tempo e qualidade de vida, o SG e o TTO são métodos mais utilizados para a obtenção das preferências para a construção dos Avaq. A aferição das preferências

por meio desses métodos envolve a realização de tarefas de difícil administração, com alto nível de abstração e requerimento cognitivo (Lamers, 2007; Robison, Loomes e Jones-Lee, 2001; Craig, 2009). A logística de campo empreendida é complexa e onerosa, na medida em que requer maior treinamento e monitoramento dos pesquisadores e maior tempo de entrevista (Stiggelbout *et al.*, 1996). O *Visual Analogue Scale* (VAS) é um método cardinal psicométrico que apresenta maior facilidade e compreensão por parte dos respondentes. Essa métrica consiste em uma escala, tipicamente representada por um termômetro vertical, com extremidades 0 e 100 (Krabbe *et al.*, 2006). O respondente posiciona cada um dos estados de saúde entre as duas extremidades, segundo suas preferências, tal que o intervalo na escala entre o valor atribuído a dois estados quaisquer corresponda à diferença entre as preferências percebidas pelo indivíduo por ambos estados de saúde. Para permitir a possibilidade de que alguns estados sejam classificados como piores que a morte, alguns sistemas de classificação nomeiam essas extremidades como melhor estado de saúde imaginável e pior estado de saúde imaginável. A classificação dos estados de saúde na escala não envolve troca entre tempo e qualidade de vida.

Métricas ordinais vêm ganhando popularidade devido a sua maior factibilidade e simplicidade na obtenção das preferências com respeito à saúde. Nessa abordagem, os métodos baseiam-se apenas na ordenação dos itens, a qual reflete as relações de preferência sem qualquer geração de números reais em uma escala intervalar, necessária para a construção dos Avaq. Um dos principais métodos ordinais é o exercício de ordenamento (Ranking). O Ranking é utilizado durante a entrevista como um procedimento que antecede a realização da VAS, TTO e SG, a fim de permitir que os respondentes se familiarizem com os estados de saúde apresentados. Esse exercício consiste em um ordenamento de estados de saúde segundo a preferência do indivíduo. O indivíduo ordena, de forma descendente, um grupo pré-determinado de estados de saúde, refletindo sobre a posição relativa que estes assumem. Recentemente, devido à simplicidade dos métodos ordinais, alguns estudos têm defendido o seu uso como método de aferição das preferências sociais (Salomon, 2003; Craig; Busschbach; Salomon, 2009; McCabe *et al.*, 2006).

Em subgrupos populacionais com baixo nível educacional ou idosos, o uso das medidas ordinais pode ser uma estratégia mais adequada para a aferição das preferências (Bansback *et al.*, 2012). Coast *et al.* (2008) mostram que a realização das tarefas envolvidas nos exercícios do TTO e SG entre os idosos é mais árdua do que no exercício de ordenação. Métodos cardinais podem causar confusão mental nos respondentes devido à necessidade de quantificação da magnitude destas preferências para diferentes estados de saúde hipotéticos, particularmente, quando se faz referência à morte. Outra questão refere-se ao alto custo das pesquisas envolvendo o 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 das métricas ordinais. O uso de abordagens relativamente mais simples pode representar uma alternativa interessante do ponto de vista dos critérios de razoabilidade das pesquisas científicas (Ali e Ronaldson, 2012).

Tanto os métodos cardinais quanto os ordinais permitem estimar preferências individuais por estados de saúde. Uma lacuna nessa literatura corresponde ao estabelecimento de qual métrica é mais adequada. Essa lacuna deve-se a, pelo menos, dois motivos. Primeiro por não existir uma medida padrão ouro para se realizar comparações entre as diferentes formas de aferir as preferências (Brazier *et al.*, 1999; McDowell, 2006). Segundo, as preferências são categorias abstratas e subjetivas tornando árdua a tarefa de identificar se o método é de fato adequado para sua mensuração. O TTO e o SG, por se basearem na teoria da Utilidade Esperada, têm sido considerados como referência para comparação com outros métodos. 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 o estabelecimento de procedimentos indiretos de se analisar essa validade. Um deles refere-se ao exame da equivalência entre os resultados obtidos de diferentes métodos (Craig, Busschbach e Salomon, 2009; Krabbe, Essink-Bot e Bonsel, 1997; Badia *et al.*, 1999). Os principais resultados de estudos de equivalência entre os métodos mostram correlação positiva e significativa entre os valores obtidos pela VAS, Ranking e pelos métodos baseados na troca. Contudo, o TTO e o SG apresentam maior número de estados de saúde classificados como piores que a morte que os demais métodos (Krabbe, Essink-Bot e Bonsel, 1997; Robison, Loomes e Jones-Lee, 2001; Brazier *et al.*, 1999; Badia, Monserrat e Herdman, 1999; Salomon, 2003; e McCabe *et al.*, 2006).

3 METODOLOGIA

3.1 O EQ-5D

Existem diversos sistemas de classificação de estados de saúde que podem ser usados na construção dos AVAQS, entre 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, esses instrumentos diferenciam-se pelo número e tipo de dimensões de saúde consideradas, além dos níveis de severidade pertencentes em cada dimensão.

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 EQ-5D é apresentado

com três níveis de severidade, EQ-5D-3L e, mais recentemente, foi desenvolvido o EQ-5D-5L, que contém cinco níveis de severidade. Até o presente momento, o EQ-5D-3L é o instrumento mais difundido e que tem sido amplamente utilizado para avaliação de tecnologias em saúde. Esse 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). O instrumento 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 de severidade cada uma (sem problemas (1), alguns problemas (2) e problemas extremos (3)), resultando na descrição de 243 (3⁵) distintos estados de saúde. Dessa forma, cada estado de saúde é representado por uma sequência de cinco algarismos, que correspondem aos níveis de severidade das respectivas dimensões. O estado de saúde 12113, por exemplo, reflete ausência de problemas na dimensão de mobilidade (1), algum problema na dimensão de autocuidado (2), ausência de problemas na realização de atividades usuais (1), ausência de dor/desconforto (1) e problemas extremos de ansiedade/depressão (3).

3.2 Base de dados

No Brasil, foram realizadas duas pesquisas para aferição de preferências sociais por estados de saúde. A primeira pesquisa foi realizada para a cidade de Porto Alegre tendo como método de aferição o SG (n = 494) (Cruz *et al.*, 2011). A segunda pesquisa foi realizada em Minas Gerais (n = 3362), em 2011, utilizando o método do TTO. O protocolo da entrevista incluiu além do TTO a aplicação da VAS e do Ranking (Andrade *et al.*, 2013). O estudo de Minas Gerais é uma oportunidade para verificar, pela primeira vez, a convergência entre as medidas de valoração de estados de saúde para uma região brasileira. Foram investigados 3.362 indivíduos alfabetizados, com idade entre 18 e 64 anos, residentes em áreas urbanas de Minas Gerais, por meio de entrevistas domiciliares. Somente um indivíduo em cada domicílio foi selecionado para compor a amostra. A amostra é probabilística, estratificada por sexo e idade e representativa para o estado, com margem de erro de 3%. O sistema descritivo EQ-5D foi traduzido para o português e foi culturalmente adaptado pelo grupo Euroqol. O protocolo de entrevistas seguiu uma versão revisada (Kind, 2009) do estudo original do MVH (Gudex, 1994). Dos 243 estados de saúde definidos pelo EQ-5D, 102 estados foram diretamente valorados. Os estados foram agrupados em 26 blocos, cada um composto por seis perfis de saúde (dois brandos, dois moderados e dois severos). A severidade dos estados de saúde é definida de acordo com sua proximidade ao melhor estado de saúde possível. Estados brandos não contêm problemas severos em qualquer dimensão; estados severos não contêm problemas brandos em qualquer dimensão;

estados moderados abrangem os demais estados que não se encaixam nas definições apresentadas (Kind, 2009). Cada indivíduo avaliou um bloco contendo seis estados de saúde e o melhor e o pior estado de saúde (11111 e 33333, respectivamente), além da “Morte”, totalizando nove estados de saúde.

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 de acordo com a Escala Analógica Visual, com as âncoras de 0 e 100 correspondendo aos piores e aos melhores estados de saúde imagináveis. Depois, por meio do Ranking, ordena-se, de forma descendente, um grupo pré-determinado e balanceado de nove estados de saúde hipotéticos definidos pelo EQ-5D, incluindo a morte. Os respondentes atribuem diferentes posições a cada estado. Ainda nesse mesmo instrumento, realiza-se o preenchimento da Visual Analogue Scale, indicando em que local da escala os participantes classificam os nove estados de saúde. Entre a aplicação do Ranking e o preenchimento da VAS, os estados de saúde são embaralhados e reapresentados de forma aleatória ao entrevistado. Ao nosso saber, essa aleatorização dos estados de saúde foi realizada pela primeira vez seguindo a proposta de Kind (2009), constituindo-se em uma primeira análise da correlação verdadeira entre as preferências geradas pelos métodos da VAS e do Ranking. Ao realizar os exercícios, os indivíduos são instruídos a se imaginarem vivendo em cada um desses estados por 10 anos sem qualquer mudança, seguido por morte.

No Instrumento B, são apresentadas aos participantes as alternativas que variam entre níveis de quantidade e qualidade de vida. Cada indivíduo avalia sete estados de saúde (seis estados que abarcam os três níveis de severidade e o pior estado de saúde hipotético 33333), utilizando o método Time Trade-off. O primeiro passo da avaliação é verificar se o indivíduo avalia cada estado de saúde como melhor ou pior do que a morte por meio de um quadro de dupla-face. No lado 1 do quadro, são avaliados os estados considerados pelos respondentes como melhores que a morte e, no outro lado, aqueles avaliados 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 esse 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 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.

3.3 Obtenção de valores para as preferências individuais por estados de saúde

Para cada um dos três métodos empregados são necessários procedimentos para estimação das preferências sociais por estados de saúde. Para o TTO e a VAS, é realizada uma transformação da variável obtida nos exercícios e estimados modelos econométricos usando os valores transformados como variável dependente para obter valores preditos para os 243 estados de saúde. No caso do ordenamento, é utilizado um modelo *logit* condicional que permite transformar variáveis ordinais em cardinais.

3.4 Time Trade-off

No TTO, a saúde integral e a morte apresentam pontuações 1 e 0 respectivamente. Estados classificados como melhores que a morte assumem valores positivos e menores do que 1 e são dados pela fórmula $V = (x/10)$, em que x é o período de tempo vivido com saúde integral. Para estados classificados como pior que a morte, os valores são dados pela fórmula $V = -x/(10-x)$. Contudo, esse procedimento gera valores com limite inferior igual a -19, resultando em uma distribuição assimétrica à esquerda. A exemplo de Dolan *et al.* (1996), a fim de tratar essa distribuição assimétrica dos valores negativos, realizou-se uma transformação monotônica por intermédio da fórmula $V_t = V/(1-V)$, se $V < 0$. Dessa forma, estados classificados como piores que a morte foram limitados inferiormente pelo valor -1, obtendo-se um intervalo equidistante da morte em ambas direções, positiva e negativa.

A partir dos 102 estados de saúde avaliados diretamente na pesquisa de campo, foi utilizado um modelo de efeitos aleatórios para dados em painel para estimar os valores das preferências sociais para todos os 243 estados de saúde. A variável dependente foi definida como 1 menos a valoração oferecida pelo TTO, $(1-V_t)$, que se refere à perda de utilidade gerada por cada condição de saúde. A especificação do modelo segue um conjunto de dez variáveis *dummy* para cada nível de severidade e dimensão de saúde conforme descrito em Andrade *et al.* (2013), que estimaram os parâmetros de valorização para MG a partir do TTO.

É importante destacar que todo o instrumento utilizado nesta pesquisa para a obtenção das preferências a partir do TTO segue o protocolo usado no estudo do *Measurement and Valuation of Health* (MVH) (Dolan, 1997), que gerou os valores por estados de saúde para o caso Inglês. O protocolo MVH foi adotado na pesquisa em Minas Gerais, pois, até o momento da realização da pesquisa de campo, era a técnica mais consolidada e amplamente utilizada para obtenção de preferências. Apesar da escolha metodológica realizada e da importância do grupo MVH para o desenvolvimento de procedimentos de pesquisa, são reconhecidos problemas conceituais e operacionais envolvidos nos instrumentos.

A principal dificuldade conceitual encontra-se na factibilidade do cenário apresentado ao respondente no que se refere aos estados classificados como piores que a morte. O indivíduo deve escolher entre morrer imediatamente ou viver por um período x de anos ($x < 10$) em uma condição específica de saúde, seguidos por $(10-x)$ anos no estado de saúde perfeita. Esse exercício requer alto nível de abstração dos respondentes, já que não parece razoável uma recuperação tão completa após um período em condição de saúde muito precária.

A inversão da ocorrência dos estados – com saúde perfeita ocorrendo anteriormente à condição precária de saúde – proposta originalmente por Torrance (1986), também impõe desafios à estimação na medida em que os indivíduos podem considerar cometerem suicídio antes do início da vigência do estado de saúde precário (Tilling *et al.*, 2008).

Outra importante fragilidade refere-se a não uniformidade entre os procedimentos para valoração de estados melhores e piores que a morte (Devlin *et al.*, 2008). Os valores dos estados considerados melhores que a morte, x/t , são obtidos com a variação de x enquanto t permanece fixo. Já no caso dos estados piores que a morte, $-x/(t-x)$, o procedimento envolve mudanças simultâneas no numerador e no denominador, o que exagera o efeito dos *trade-offs* realizados. No âmbito operacional, a dificuldade encontra-se no fato de o protocolo requerer que o participante se engaje em tarefas totalmente diferentes, com o uso de abordagens e instrumentos visuais distintos para a obtenção de uma mesma distribuição de valores. Essa descontinuidade entre as tarefas e valores para estados melhores e piores que a morte gera um *gap effect* na estimação das utilidades, já que poucos estados se localizam próximos à morte (zero) (Stalmeier *et al.*, 2005; Craig e Busschbach, 2009). Considerando todos os problemas envolvidos nesses procedimentos, a validade da agregação de valores positivos e negativos para um dado estado de saúde a fim de se gerar um conjunto de utilidades esperadas parece, no mínimo, questionável. A geração de valores negativos com limites superiores à unidade e ao tratamento dado em questão também são pontos centrais de discussão. Dada a inexistência de um fundamento teórico para a imposição de limites do nível de desutilidade associada a situações de extremo sofrimento, valores negativos possuem maior peso no cálculo das utilidades médias. Para evitar esse desequilíbrio, os valores negativos usualmente são limitados em -1. Lamers (2007) mostra como diferentes métodos para limitar valores negativos afetam a estimação do conjunto de preferências. As transformações realizadas ocasionam a elevação dos valores para todos os estados de saúde. Além disso, as diferenças em relação aos dados brutos aumentam à medida que a condição de saúde se torna mais severa.

Alguns estudos apresentam alternativas interessantes ao uso do procedimento padrão de valoração para estados de saúde piores que a morte. Empregando utilidades

sem quaisquer transformações, Craig e Busschbach (2009) mesclam as abordagens TTO e *Discrete Choice* em um único estimador, aumentando a validade de convergência entre métodos psicométricos e econômicos. Já Robinson e Spencer (2006) desenvolveram uma abordagem que pode ser aplicada uniformemente para ambos os estados, piores e melhores que a morte. A principal inovação do método é a introdução de um tempo antecedente adicional (*lead time*) em saúde perfeita para cada estado a ser avaliado. Assim, retira-se a necessidade de diferentes procedimentos em um mesmo método de valoração ao permitir que os participantes troquem seu tempo adicional a fim de evitarem condições de saúde muito precárias. O único impacto esperado com a introdução do *lead time* é a forma com que os estados de saúde piores que a morte são avaliados. Devlin *et al.* (2008) tem considerado essa abordagem a alternativa mais promissora para a resolução dos problemas apresentados no procedimento padrão. Como esperado, o método supera a descontinuidade em torno do valor zero apresentada pela abordagem MVH. Contudo, os autores destacam algumas dificuldades teóricas do método e a necessidade de maiores investigações para entender suas implicações. Há indícios de que a introdução do *lead time* possui efeitos complexos sobre a distribuição das observações positivas relacionadas aos estados melhores que a morte. Além disso, o método reforça o efeito duração (Craig, 2009) presente no protocolo original, já que o valor obtido para um estado de saúde de interesse pode ser afetado pela duração do tempo antecedente adicional em saúde perfeita que é oferecido ao participante. Se assumirmos que os participantes empregam desconto intertemporal no exercício TTO, o valor marginal do tempo gasto em determinado estado de saúde pode não ser constante.

3.5 Visual Analogue Scale (VAS)

As pontuações obtidas diretamente a partir da VAS também necessitam ser transformadas para que os estados de saúde perfeita e morte representem âncoras para todos os respondentes. As pontuações obtidas na VAS são transformadas

segundo a fórmula $VAS_j^t = \frac{VAS_j - VAS_{MORTE}}{VAS_{PERFEITA} - VAS_{MORTE}}$. A estimação dos valores para os

243 estados de saúde é realizada utilizando o modelo de efeitos aleatórios para dados em painel, similar ao implementado para o TTO. A especificação do modelo segue a mesma abordagem, tendo como variáveis independentes um conjunto de dez variáveis *dummy* definidas para cada dimensão e nível de severidade.

3.6 Ranking

A análise dos dados ordinais é baseada em McFadden (1974), que propôs um modelo de utilidade aleatória como meio de estimar a função de utilidade latente. Esse modelo é operacionalizado por meio do modelo *logit* condicional. Neste trabalho, a estimação de preferências obtidas a partir de métodos ordinais segue a proposta

por Salomon (2003), McCabe *et al.* (2006), usualmente empregada na literatura (Craig, Busschabach e Salomon (2009), Brazier *et al.* (2009) e Ryan *et al.* (2006)). O método *logit* condicional pressupõe que os indivíduos, ao realizarem o exercício de ordenação, estabelecem uma série de seleções em conjuntos de estados de saúde cada vez menores. Em uma ordenação de n estados de saúde, 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 . O respondente i atribui um valor de utilidade latente para o estado j , U_{ij} , que inclui um componente sistemático e um termo de erro:

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

O componente sistemático é determinado somente por características do estado de saúde, sendo μ , portanto, indexado somente por j . Dessa forma, um dado estado de saúde tem um mesmo valor esperado de utilidade sombra entre todos os respondentes. As posições observadas para um conjunto de estados de saúde são supostas correlacionadas a esses valores, sendo distribuídos em torno de um nível médio para cada estado. Um indivíduo somente atribui 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; McCabe *et al.*, 2006). Ao escolher o estado preferível em um grupo finito de estados, o indivíduo i ordenará o estado j acima do estado k se $U_{ij} > U_{ik}$, para todo $j \neq k$. Logo, $(\mu_j - \mu_k) > (\varepsilon_k - \varepsilon_j)$. O modelo de utilidade aleatória adota a hipótese de independência das alternativas irrelevantes (IIA). De acordo com essa hipótese, 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 . Para verificar a validade da IIA, utilizou-se o teste proposto por Hausman e McFadden (1984), o qual consiste da análise da consistência dos parâmetros estimados utilizando apenas um subgrupo de alternativas. Para tanto, foi excluída, aleatoriamente, uma alternativa para cada respondente. O modelo é re-estimado usando a amostra reduzida, comparando os coeficientes estimados com aqueles da amostra completa.

No contexto do exercício de ordenação, μ_j pode ser entendido como a valoração média de um estado de saúde, a qual pode ser expressa como função das dimensões e níveis do sistema descritivo EQ-5D. O valor esperado da utilidade latente de cada estado de saúde é uma função linear dos cinco domínios de saúde. A variável dependente refere-se à posição dada por cada indivíduo aos estados de saúde Ranking. A exemplo do modelo de efeitos aleatórios utilizado para a análise das informações obtidas pelo TTO (Andrade *et al.*, 2013), define-se θ o 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. Essas variáveis são definidas para cada um dos estados de saúde hipotéticos avaliados pelos indivíduos. Além disso, inclui-se uma variável categórica, λ , que toma valor igual a 1, se o estado é igual à Morte e 0, caso contrário. O valor de μ_j corresponde ao valor médio dos estados de saúde, predito ou recomposto a partir dos coeficientes estimados e das variáveis *dummy*.

$$\mu_j = \theta x_j + \lambda D. \quad (2)$$

Os modelos são estimados considerando a correção para observações não independentes e identicamente distribuídas (iid), uma vez que cada indivíduo avalia nove estados de saúde. Os indivíduos que atribuíram a mesma posição para dois estados de saúde tiveram o ordenamento duplicado na estimação da máxima verossimilhança. Cada ordenamento recebeu uma ponderação de 50% diferindo apenas na posição dos estados de saúde originalmente classificados como indiferentes (Craig; Busschabach; Salomon, 2009).

3.7 Normalização e medidas de comparação

Para obter valores para as preferências sociais por estados de saúde, é necessário estabelecer uma ancoragem para a distribuição dos valores relativos obtidos no modelo *logit* condicional. No método TTO, a tarefa de valoração realizada pelos indivíduos pressupõe os estados de saúde perfeita e morte como âncoras. Dessa forma a morte recebe valor 0 e a saúde perfeita valor igual a 1, respeitando a distribuição de valores do Avaq. Neste trabalho a ancoragem adotada segue o mesmo procedimento do TTO (Salomon, 2003; McCabe *et al.*, 2006; Brazier *et al.*, 2009). A ordem em que os estados de saúde são classificados mantém-se diante de uma transformação afim positiva das utilidades latentes, implicando a seguinte especificação geral da equação (1):

$$U_{ij} = \alpha(\mu_j + \varepsilon_{ij}) + \beta. \quad (3)$$

Substituindo na equação (3) a utilidade predita para um determinado estado de saúde (2) condicional aos valores dos parâmetros estimados no modelo, tem-se:

$$U_{ij} = \alpha(\theta x_j + \lambda D) + \beta. \quad (4)$$

O parâmetro β representa o valor do estado de saúde perfeita (11111). Assim, quando um perfil não possui dificuldade em nenhuma das dimensões ($x_j = 0$) e não se refere à condição de morte ($D = 0$), o estado de saúde terá valor esperado

U_{ij} igual a 1 e, portanto, normaliza-se $\beta = 1$. Na condição de morte, x_j é igual a 0 e D é igual a 1. Assim, para atribuir utilidade zero ($U_{ij} = 0$) à condição de morte, α é determinado por:

$$\alpha \hat{\lambda} + 1 = 0. \tag{5}$$

$$\alpha = -\frac{1}{\hat{\lambda}}. \tag{6}$$

Dessa forma, mantêm-se as âncoras saúde perfeita e morte com valores 1 e zero, respectivamente. Esse procedimento possibilita que estados de saúde classificados como piores que a morte apresentem valores negativos.

Após a estimação das preferências pelos três métodos, procede-se à comparação das distribuições desses valores. A comparação é realizada entre os valores preditos da VAS e do Ranking em relação aos valores estimados no método TTO. Essa comparação pressupõe que o TTO é o método padrão empregado para estimação de preferências sociais por estados de saúde. A análise da distribuição dos valores preditos, segundo os três métodos, é realizada utilizando os coeficientes de *Spearman*, *Pearson* e *Kendall Tau*. Enquanto o primeiro é uma medida de correlação não paramétrica, o segundo requer a suposição de que a relação entre as variáveis seja linear. A correlação de Kendall é uma medida de associação para variáveis ordinais e representa a divergência entre a verdadeira concordância mostrada nos dados e a concordância perfeita ou máxima possível (Siegel e Castellan Jr., 2006). Nesse sentido, os valores preditos para os 243 estados de saúde a partir dos três métodos de valoração são transformados em sequências ordinais a fim de se analisar este grau de associação. Os valores estimados a partir do Ranking e da VAS são comparados aos valores preditos do TTO, calculando-se a estatística da Diferença Absoluta Média (DAM). O ajuste dos métodos é mensurado usando a DAM e o número de estados de saúde com diferenças absolutas superiores a 0,05 e 0,10. As análises estatísticas são computadas usando o software Stata 12.0.

4 RESULTADOS

A tabela 1 apresenta os resultados do modelo *logit* condicional para o Ranking e do modelo de efeitos aleatórios para o TTO e a VAS. Todos os coeficientes são positivos e significativos a 1%. Assim como os coeficientes do TTO e da VAS, o modelo *logit* condicional apresenta comportamento esperado quanto aos sinais dos parâmetros estimados. No modelo de efeitos aleatórios, verifica-se

um aumento monotônico no valor da desutilidade à medida que se eleva o nível de severidade em todas as dimensões. O mesmo comportamento é observado nos coeficientes estimados a partir do método ordinal referente ao método Ranking. O maior decréscimo de utilidade foi observado para problemas severos de mobilidade, enquanto o menor foi observado para problemas moderados no domínio ansiedade/depressão. Portanto, verifica-se equivalência entre os modelos no que se refere à importância relativa das dimensões e níveis de severidade.

TABELA 1
Resultados do modelo *logit* condicional (Ranking) e do modelo de efeitos aleatórios (Time Trade-off e Visual Analogue Scale)

Variáveis	TTO	d. p.	VAS	d. p.	Ranking	d. p.
Mobilidade 2	0,127***	0,007	0,150***	0,006	1,140***	0,029
Cuidados pessoais 2	0,121***	0,007	0,136***	0,006	0,686***	0,026
Atividades habituais 2	0,095***	0,007	0,108***	0,007	0,512***	0,026
Dor/desconforto 2	0,067***	0,007	0,079***	0,006	0,435***	0,022
Ansiedade/depressão 2	0,061***	0,007	0,065***	0,007	0,406***	0,025
Mobilidade 3	0,403***	0,008	0,259***	0,008	2,261***	0,045
Cuidados pessoais 3	0,246***	0,008	0,171***	0,007	1,029***	0,031
Atividades habituais 3	0,205***	0,008	0,149***	0,007	0,859***	0,031
Dor/desconforto 3	0,200***	0,007	0,118***	0,007	0,729***	0,026
Ansiedade/depressão 3	0,113***	0,007	0,087***	0,007	0,651***	0,027
Morte	-	-	-	-	7,373***	0,091
Número de observações	23.300		26.736		30.105	

Elaboração dos autores.

Nota: *** Significante a 1%.

O teste da hipótese de independência das alternativas irrelevantes do modelo *logit* condicional é reportado na tabela 2. A partir da análise do *p-valor* do teste de *Hausman*, não se pode rejeitar a hipótese nula de igualdade entre os coeficientes da amostra completa e da amostra reduzida que exclui aleatoriamente uma observação de cada respondente, sugerindo que o modelo *logit* condicional pode ser utilizado na estimação das preferências dos estados de saúde com base no método Ranking.

TABELA 2
Teste de *Hausman* para a hipótese de independência das alternativas irrelevantes

Variáveis	Amostra completa	d. p.	Amostra reduzida	d. p.
Mobilidade 2	1,14	0,028	1,146	0,03
Cuidados pessoais 2	0,686	0,025	0,692	0,028
Atividades habituais 2	0,512	0,027	0,517	0,029
Dor/desconforto 2	0,435	0,023	0,431	0,025
Ansiedade/depressão 2	0,406	0,025	0,415	0,028

(Continua)

(Continuação)

Variáveis	Amostra completa	d. p.	Amostra reduzida	d. p.
Mobilidade 3	2,261	0,034	2,279	0,037
Cuidados pessoais 3	1,029	0,028	1,024	0,031
Atividades habituais 3	0,859	0,029	0,868	0,031
Dor/desconforto 3	0,729	0,026	0,729	0,029
Ansiedade/depressão 3	0,651	0,026	0,649	0,029
Morte	7,373	0,055	7,363	0,061
Teste Hausman	$\chi^2 = 8,12$		p-valor = 0,703	

Elaboração dos autores.

Obs.: Todos os valores são significativos a 1%.

A tabela 3 apresenta os resultados da análise de correlação entre os valores preditos obtidos com os três métodos de valoração. As estatísticas de Pearson e Spearman indicam que os valores preditos são altamente correlacionados, com coeficientes de correlação acima de 0,94 (tabela 3).

TABELA 3
Correlações de Pearson e Spearman dos valores esperados por método

Correlação de Pearson			
	TTO Predito	VAS Predito	Ranking
TTO Predito	1,000	-	-
VAS Predito	0,946	1,000	-
Ranking	0,966	0,975	1,000
Correlação de Spearman			
	TTO Predito	VAS Predito	Ranking
TTO Predito	1,000	-	-
VAS Predito	0,976	1,000	-
Ranking	0,981	0,989	1,000

Elaboração dos autores.

Obs.: Todos os valores são significativos a 1%.

A análise da concordância das ordenações dos estados de saúde geradas a partir dos três métodos revela paridade entre os valores. A correlação Tau de Kendall é igual a 0,799, entre TTO e VAS, e a 0,842 entre TTO e Ranking. Esse resultado evidencia que a ordenação realizada por meio dos três métodos é muito próxima. Valores próximos de 1 indicam elevado grau de concordância entre todos os pares comparados.

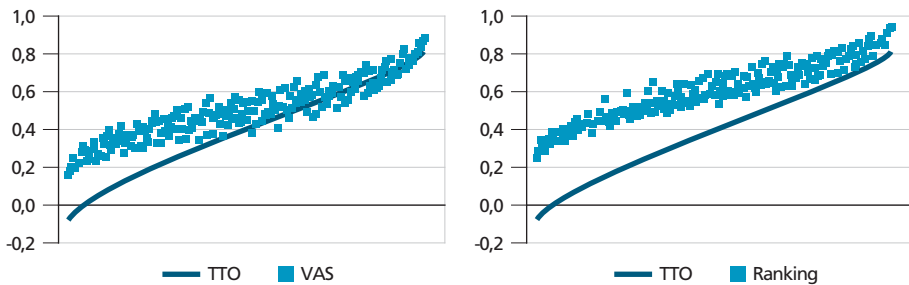
A similaridade entre as ordenações dos estados de saúde não garante similaridade na cardinalidade estimada para as preferências. O gráfico 1 ilustra o ajuste entre os valores preditos obtidos por meio dos três métodos para as preferencias sociais por estados de saúde. A análise do gráfico 1 mostra disparidade principalmente entre os valores estimados para os estados de saúde mais severos. As diferenças

absolutas médias entre VAS e TTO e entre Ranking e TTO são iguais a 0,117 e 0,208, respectivamente. Como o valor absoluto estimado para a preferência situa-se entre 0 e 1, esse resultado significa que, na média, os valores preditos da VAS e do Ranking distam em cerca de 12% e 21%, respectivamente, dos valores estimados no TTO.

A análise por grupos de estados de saúde mostra que essa diferença depende do nível de severidade. Na VAS, a DAM apresenta valores iguais a 0,028, 0,117 e 0,186 para estados classificados como brandos, moderados e severos, respectivamente. No Ranking, esses valores são iguais a 0,073, 0,211 e 0,307. Esse fato pode estar 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. Isso sugere que os indivíduos podem estar menos propensos a classificar estados de saúde como piores que a morte nos métodos Ranking e VAS. A posição ocupada pelo cartão Morte nos diferentes métodos de valoração corrobora esse resultado. Cerca de 11,7% e 10,6% dos respondentes classificam algum estado de saúde como pior que a morte no Ranking e na VAS, enquanto no TTO, esse percentual é de quase 73,5%. No que se refere ao número de respostas, o TTO também apresenta maior percentual de *trade-off* entre qualidade e quantidade. Do total de 30.258 observações, 2,6% (799), 2,8% (849) e 21% (6423) representam valorações piores que a morte.

GRÁFICO 1

Valores preditos para 243 estados de saúde a partir dos métodos TTO, VAS e Ranking

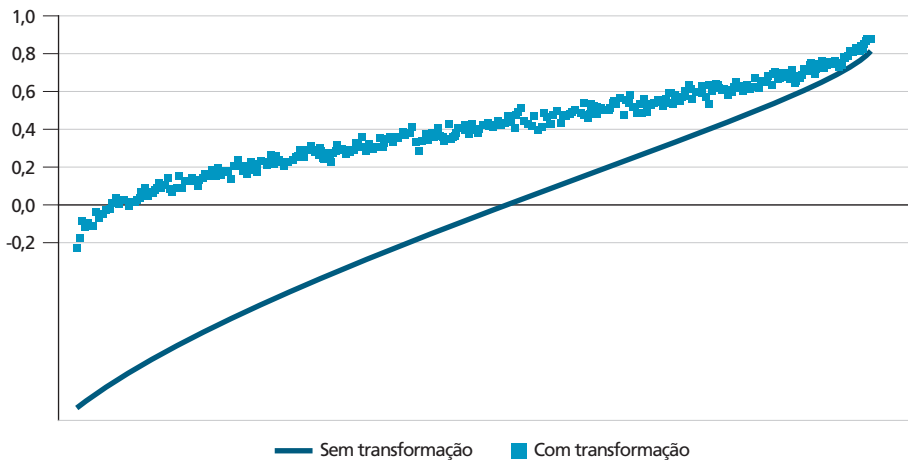


Elaboração dos autores.

A adoção do método de transformação dos valores associados aos estados piores que a morte, no TTO, parece afetar a estimação das utilidades. O gráfico 2 compara os valores preditos de acordo com as informações brutas captadas pelo método e as utilidades estimadas após a transformação monotônica dos valores negativos. Apesar de a ordinalidade das preferências manter-se, as diferenças cardinais são expressivas. Observa-se que a estimação realizada com base nos dados brutos produz diminuição das médias dos estados de saúde, com uma gama maior de perfis

classificados como piores que a morte, inclusive com alguns poucos valores negativos fora do raio unitário. Todos os seis estados de saúde com valores inferiores a -1 possuem pelo menos quatro dimensões com nível de severidade extrema (32333, 33233, 33323, 33331, 33332, 33333). A introdução da transformação de valores reduziu de 55% (132) para 5% (12) o número de estados de saúde classificados como piores que a morte. No que se refere à comparação de métodos, a diferença entre os valores estimados com e sem transformação monotônica faz com que as utilidades esperadas do TTO se distanciem ainda mais daquelas encontradas na VAS e no Ranking. As diferenças absolutas médias entre VAS e TTO e entre Ranking e TTO passam a ser 0,589 e 0,689, respectivamente.

GRÁFICO 2
Valores preditos a partir do TTO com e sem transformação para WTD



Elaboração dos autores.

5 DISCUSSÃO

Este trabalho tem como objetivo verificar a equivalência de três métodos de aferição de preferências por estados de saúde – Ranking, Visual Analogue Scale e Time Trade-off – e discutir como esses resultados podem afetar o processo decisório de tratamentos e programas de saúde no âmbito do SUS. Os resultados sugerem que apesar da ordenação das preferências por estados de saúde ser bastante similar entre os três métodos analisados – corroborando estudos anteriores – diferenças expressivas de magnitude são observadas. Apesar das altas correlações encontradas entre os diferentes métodos, as valorações dos estados diferiram de forma significativa. A diferença absoluta média entre os valores observados no TTO e no Ranking foi superior à encontrada em grande parte dos estudos internacionais. A diferença mostra-se expressiva para os estados mais 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.

Badia, Monserrat e Herdman (1999) encontraram alta correlação entre os métodos, apesar de existirem diferenças entre eles no que se refere ao número de estados classificados como piores que a morte. Os valores do TTO são maiores que aqueles encontrados pela VAS para estados de saúde brandos, mas consideravelmente menores para estados mais severos. Segundo Robison, Loomes e Jones-Lee (2001), a diferença entre os valores obtidos a partir do VAS e do TTO é um reflexo da diferença entre os pontos de referência adotados pelo respondente em cada um dos métodos. Na VAS, o indivíduo toma sua própria condição de saúde como referência para avaliar os demais estados. No TTO, os pontos de referência são a morte e a saúde perfeita. Assim, os processos cognitivos que os respondentes utilizam na realização das tarefas são qualitativamente distintos. Salomon (2003) e McCabe *et al.* (2006) encontram grande equivalência entre os valores estimados no Ranking e no TTO para o Reino Unido. Resultados similares são encontrados por Craig, Busschbach e Salomon (2009), examinando oito países: Eslovênia, Argentina, Dinamarca, Japão, Holanda, Espanha, Reino Unido e Estados Unidos. Contudo, o Ranking gera menor proporção de estados de saúde piores que a morte que o TTO. Utilizando dois sistemas descritivos de saúde específicos, Brazier *et al.* (2012) encontram no Ranking coeficientes similares aos encontrados pelo TTO além da correlação entre os valores obtidos a partir dos dois métodos ser igual a 0,715.

Em termos de política pública, os resultados encontrados no estudo para Minas Gerais mostram que os métodos Ranking e VAS podem ser utilizados para ordenação de diferentes tecnologias em saúde. A ordenação dos benefícios (medidos pelo AVAQ) associados às diferentes tecnologias em saúde será a mesma, independentemente, do método de valoração utilizado. A limitação do uso desses métodos decorre das diferenças encontradas na cardinalidade. Se a introdução de uma intervenção ou tecnologia no sistema de saúde estiver determinada pela definição de um valor limite (*threshold*) para a razão de custo efetividade, as diferenças de magnitude entre os três métodos poderão determinar decisões distintas. No Brasil, desconhece-se a existência de um *guideline* referente a um *limiar de aceitabilidade* para análises de custo-efetividade. Um *threshold* de £ 20-30.000 parece ser amplamente utilizado no caso inglês. À medida que aumenta a razão custo por AVAQ ganho, diminui-se a probabilidade de se adotar uma dada tecnologia (McDonough e Tosteson, 2007). A Organização Mundial da Saúde (OMS) sugere um valor limite igual a três vezes o produto interno bruto *per capita* do país onde se concentra a análise (Bertoldi *et al.*, 2013; Rodrigues *et al.*, 2013).

A fim de se ilustrar o impacto dos achados deste trabalho sobre as decisões no âmbito do sistema público de saúde, suponha a avaliação de um dado tratamento hipotético para pacientes acometidos por artrite reumatoide. Para realizar essa simulação, além dos pesos médios por estados de saúde estimados para Minas Gerais, por meio dos três métodos de valoração, utilizamos as informações referentes aos perfis medianos de saúde de pacientes diagnosticados com a doença, segundo descrição do EQ-5D (Hurst *et al.*, 1997). Usualmente, os pacientes de artrite reumatoide são estratificados em classes funcionais que representam os diferentes níveis de severidade da doença (Hochberg *et al.*, 1992). Se assumirmos que, com o tratamento, os pacientes conseguem evoluir de uma condição severa de saúde (Classe IVa) para um quadro geral mais brando (Classe I), conseguimos estimar o ganho da intervenção hipotética, em termos de anos de vida ajustados pela qualidade – Avaq. Supondo um ganho em anos de vida igual a 1 e um custo incremental anual médio hipotético de R\$ 29.000 (valor corrente de 2012) – igual ao gasto médio/paciente com os principais medicamentos ofertados pelo sistema público de saúde brasileiro para o tratamento de artrite (Brasil, 2012). Os resultados das razões de custo-efetividade incremental dependerão do método de obtenção de preferências utilizado. Enquanto o TTO produz um custo de R\$ 53.211 por Avaq ganho, na VAS e no Ranking essas estimativas chegam a R\$ 67.757 e R\$ 73.047, respectivamente (apêndice A). Se adotarmos o valor limite igual a três vezes o PIB *per capita*, sugerido pela OMS para a introdução de uma dada tecnologia, a razão Custo/Avaq para intervenções em saúde no Brasil não deve ultrapassar R\$ 66.132 (PIB *per capita* brasileiro corrente de 2012 igual a R\$ 22.044) (IBGE, 2015). Dessa forma, se as análises custo-utilidade fossem realizadas baseadas apenas a partir do uso da VAS ou Ranking, o dado tratamento para pacientes com artrite reumatoide não seria introduzido na rede SUS.

Cabe destacar que o TTO é o método dominante no mundo para aferição das preferências por estados de saúde. Contudo, tem sido criticado por requerer elevada capacidade cognitiva, pela dificuldade de administração do protocolo da entrevista e pela necessidade de transformação dos valores dos estados classificados como piores que a morte (Craig, Busschbach e Salomon, 2009). De fato, como mostrado neste estudo, transformações de valores negativos associados a estados de saúde piores que a morte, tem o potencial de afetar as utilidades médias estimadas. Em países com maior heterogeneidade socioeconômica e menor nível de escolaridade, as limitações envolvidas no uso do método TTO podem ser mais acentuadas. A heterogeneidade cultural e socioeconômica é presente na América Latina, principalmente no Brasil. Enquanto no Canadá, Estados Unidos e Alemanha somente 2,4%, 1,7% e 5,0%, respectivamente, da população com idade superior a 15 anos possuem apenas o ensino primário completo, no Chile, Argentina e Brasil, essas estimativas chegam a 12,3%, 32,7% e 27,3%, em 2010 (The World Bank, 2014).

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. A recente difusão da avaliação de tecnologias em saúde em países em desenvolvimento torna relevante analisar em que medida métodos de valoração mais simples – como VAS e Rank – geram preferências similares às aquelas obtidas através do TTO. Nesse sentido, este trabalho contribui para esta discussão ao revelar a viabilidade e alguns desafios no uso de técnicas mais simples de obtenção de preferências por estados de saúde.

O marco institucional para definição de introdução de tecnologias em saúde no Brasil tem se consolidado nos últimos anos principalmente a partir de 2004 com a criação do Departamento de Ciência e Tecnologia (Decit). O Decit é responsável pela formulação e promoção da avaliação de tecnologias em saúde para o Sistema Único de Saúde (SUS). Mais recentemente, em 2011, foi criada a Comissão Nacional de Incorporação de Tecnologias no SUS (Conitec). De acordo com a Lei Federal nº 12.401/2011, a incorporação, exclusão ou alteração de novos medicamentos, produtos e procedimentos que serão fornecidos no SUS devem ser avaliadas pelo Conitec. Esse é um grande avanço da legislação brasileira, uma vez que parâmetros de custo-efetividade são agora considerados na decisão de incorporação de novas tecnologias. Um desafio para essa Comissão é o de incluir nas avaliações de tecnologias em saúde medidas de qualidade de vida. Os ganhos de longevidade podem não ser uma métrica suficiente para analisar os ganhos de bem-estar no que se refere ao estado de saúde. A realização de ATS deverá considerar as especificidades do processo de transição demográfica e epidemiológica e as grandes disparidades socioeconômicas e regionais presentes no país. Por um lado, nas regiões mais desenvolvidas, como o Sul e Sudeste, por apresentar uma população mais envelhecida, o perfil epidemiológico é caracterizado pelo predomínio de doenças crônicas. Essas doenças demandam cuidados prolongados, com maior conteúdo tecnológico e uso constante de medicamentos. 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 elevadas de mortalidade, sobretudo infantil. Portanto, as avaliações econômicas em saúde realizadas no país frequentemente deparam-se com tecnologias que são direcionadas à prevenção e ao tratamento de doenças com diferentes níveis de letalidade e morbidade. Avaliações econômicas baseadas na análise custo-utilidade, ao combinar informações sobre mortalidade com uma medida de qualidade de vida relacionada à saúde, permitem contemplar os diferentes

perfis epidemiológicos que existem no Brasil. Nesse contexto, discussão sobre qual a melhor métrica para se medir os benefícios associados a diferentes intervenções de saúde será cada vez mais crucial para se promover alocações mais eficientes dos recursos. Devido à escassez de recursos (Servo *et al.*, 2011) e à crescente intervenção do poder judiciário no setor (Borges e Ugá, 2009), a expectativa de obtenção de um sistema capaz de garantir uma assistência integral à população representa um grande desafio aos gestores. Nas próximas décadas, as políticas públicas na área da saúde possivelmente serão influenciadas pelos marcos legais que entraram em vigor nos últimos anos, de modo a permitir avanços nas decisões de alocação de recursos, especialmente, no âmbito de um sistema constitucionalmente universal, como é o caso brasileiro.

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, n. 1, p. 21-44, 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.
- 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, n. 4, 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.* Using a discrete choice experiment to estimate health state utility values. **Journal of Health Economics**, v. 31, n. 1, p. 306-318, Jan. 2012.
- BERTOLDI, E. *et al.* Cost-effectiveness of cardiac resynchronization therapy in patients with heart failure: the perspective of a middle-income country's public health system. **International Journal of Cardiology**, v. 163, n. 3, p. 309-315, 2013.
- BORGES, D. C. L.; UGÁ, M. A. D. As Ações individuais para o fornecimento de medicamentos no âmbito do SUS: características dos conflitos e limites para a atuação judicial. **Revista de Direito Sanitário**, São Paulo, v. 10, n. 1, p. 13-38, 2009.
- 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.

_____. _____. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. **Medicamentos Biológicos** (infiximabe, etanercepte, adalimumabe, rituximabe, abatacepte, tocilizumabe, golimumabe e certolizumabe pegol) para tratamento da Artrite Reumatoide. Brasília, 2012.

_____. Ministério da Saúde. Departamento de Ciência e Tecnologia. **Sisrebrats** – Sistema de Informação da Rede Brasileira de Avaliação de Tecnologia em Saúde. Disponível em: <goo.gl/HVw90N>. Acesso em: 7 maio 2015.

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. *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. **The European Journal of Health Economics**, v. 13, n. 5, p. 575-587, Oct. 2012.

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**, 2009. (Discussion Paper Series n. 9).

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

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. The duration effect: a link between TTO and VAS values. **Health Economics**, v. 18, n. 2, p. 217-225, 2009.

CRAIG, B; BUSSCHBACH, J. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. **Population Health Metrics**, v. 7, n. 3, p. 1-10, 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, n. 3, p. 296-305, 2009.

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.

DEVLIN, N. *et al.* **A uniform time trade off method for states better and worse than dead**: feasibility study of the 'lead time' approach. School of Social Sciences, Department of Economics, City University London, UK. 2008. (Discussion Paper Series n. 9).

DOLAN, P. *et al.* Valuing health states: a comparison of methods. **Journal of Health Economics**, v. 15, n. 2, p. 209-231, 1996.

DOLAN, P; SUTTON, M. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. **Soc Sci Med**, v. 44, n. 10, p. 1519-1530, 1997.

DOLAN, P. Whose Preferences Count? **Med Decis Making**, v. 19, n. 4, p. 482-486, 1999.

DOLAN, P. Modelong valuations for EuroQol health states. **Medical Care**, v. 35, n. 11, p. 1095-108, 1997.

DRUMMOND, M; SCULPHER, M. Common methodological flaws in economic evaluations. **Medical Care**, v. 43, Supp. 7, p. 5-14, 2005.

DRUMMOND, M.; STODDART, G.; TORRANCE, G. **Methods for the economic evaluation of health care programmes**. 2. ed. Oxford: Oxford Medical Publications, 1997.

FURLONG, *et al.* **Guide to design and development of health-state utility instrumentation**. Hamilton, Ontário, Canada: McMaster University, 1990. (Working Paper Series n. 90-0).

GABRIEL, S. *et al.* Health-related Quality of Life in Economic Evaluations for Osteoporosis: Whose Values Should We Use? **Med Decis Making**, v. 19, n. 2, p. 141-148, 1999.

GUDEX, C. **Time trade-off user manual**: props and self-completion method. Centre for Health Economics, University of York, Occasional Paper Series, 1994.

HAUSMAN, J.; MCFADDEN, D. Specification tests for the multinomial logit model. **Econometrica**, v. 52, n. 5, p. 1219-1240, 1984.

HOCHBERG, M. *et al.* The american college of rheumatology 1991: revised criteria for the classification of global functional status in rheumatology arthritis. **Arthritis & Rheumatism**, v. 35, n. 5, p. 498-502, 1992.

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.

HURST, N. *et al.* Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of euroqol (EQ-5D). **British Journal of Rheumatology**, v. 36, n. 5, p. 551-559, 1997.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Brasil em Síntese**. Contas Nacionais, 2015.

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

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.

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, 2006.

KRABBE, P.; ESSINK-BOT, M.; BONSEL, G. The Comparability and Reliability of Five Health-State Valuation Methods. **Social Science & Medicine Journal**, v. 45, n. 11, p. 1641-1652, 1997.

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, 2007.

LENERT, L. *et al.* The effect of search procedures on utility elicitations. **Med Decis Making**, v. 18, n. 1, p. 76-83, 1998.

LUCE, B. What will it take to make cost-effectiveness analysis acceptable in the United States? **Medical Care**, v. 43, Supp. 7, p. 44-48, 2005.

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

MCDONOUGH, C; TOSTESON, A. Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. **Pharmacoeconomics**, v. 25, n. 2, p. 93-106, 2007.

MCDOWELL, I. **Measuring health**: a guide to rating scales and questionnaires. 3. 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, p. 105-142, 1974.

MCTAGGART-COWAN, H. *et al.* The impact of disease adaptation information on general population values for rheumatoid arthritis states. **Health Economics and Decision Science**. University of Sheffield, 2010. (Discussion Paper n. 10).

NICE – NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE. **Guide to the methods of technology appraisal**. Nice, 2008.

NEUMANN, P. *et al.* Growth and quality of the cost-utility literature, 1976-2001. **Value in Health**, v. 8, n. 1, p. 3-9, 2005.

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

PHILLIPS, C. What is a QALY? **Health Economics**. What is...? Series, 2. ed. UK: Hayward, 2009.

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, 2009.

ROBINSON, A.; SPENCER, A. Exploring challenges to TTO utilities: valuing states worse than dead. **Health Economics**, v. 15, n. 4, p. 393-402, 2006.

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, 2001.

RODRIGUES, M. *et al.* Custo-efetividade da inclusão dos respondedores virológicos lentos no tratamento da hepatite C na presença da coinfeção com o HIV. **Cadernos de Saúde Pública**, v. 29, Supl. 1, p. S146-S158, 2013.

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, 2006.

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, n. 1, p. 1-12, 2003.

SERVO, L. M. S *et al.* Financiamento e Gasto Público de Saúde: Histórico e Tendências. *In*: MELAMED, C.; PIOLA, S. F. (Orgs.). **Políticas públicas e financiamento federal do Sistema Único de Saúde**. Brasília: Ipea, 2011. 356 p.

SIEGEL, S.; CASTELLAN Jr., N. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006. 448 p.

STALMEIER, P. *et al.* The gap effect: discontinuities of preferences around dead. **Health Economics**, v. 14, n. 7, p. 679- 685, 2005.

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.

THE WORLD BANK. World Data Bank. **Education Statistics**. Disponível em: <goo.gl/hl7Fqh>. Acesso em: 30 set. 2014.

TILLING, C. *et al.* Protocols for TTO valuations of health states worse than dead: a systematic review and framework for analysis. **Health Economics and Decision Sciences**, School of Health and Related research, University of Sheffield, 2008. (Discussion Paper Series n. 8).

TORRANCE, G. Measurement of health state utilities for economic appraisal: a review. **Journal of Health Economics**, v. 5, n. 1, p. 1-30, 1986.

APÊNDICE

TABELA A.1

Razão custo-utilidade para tratamento hipotético de artrite reumatoide, segundo métodos de obtenção de preferências por estados de saúde

	23322 ¹	11221 ¹	Δ AVAQ	Custo ² /AVAQ
TTO	0,238	0,783	0,545	R\$ 53.211
VAS	0,331	0,759	0,428	R\$ 67.757
Ranking	0,475	0,872	0,397	R\$ 73.047

Elaboração dos autores.

Notas: ¹ Valores preditos para os respectivos estados de saúde com base nas preferências estimadas para Minas Gerais.

² Custo do tratamento igual a R\$ 29.000.