

O sistema português de valores do SF-6D

PEDRO LOPES FERREIRA
LARA NORONHA FERREIRA
LUÍS NOBRE PEREIRA

Introdução: Nos últimos tempos tem-se assistido ao desenvolvimento de estudos em alguns países, de forma a obter funções de valoração relativas aos instrumentos de medição de preferências mais utilizados. Na verdade, a evidência sugere que as valorações dos estados de saúde podem diferir de país para país. O objectivo deste estudo foi a determinação de um sistema de valores subjacente ao SF-6D para Portugal.

Metodologia: Uma amostra de 55 estados de saúde definidos pelo SF-6D foi valorizada por uma amostra aleatória estratificada da população em geral, utilizando o *standard gamble* (SG). A pesquisa foi conduzida por entrevistas pessoais, tendo os respondentes valorizado seis estados de saúde, acrescidos do pior estado de saúde definido pelo SF-6D ou da morte imediata. Estimaram-se vários modelos

com o objectivo de prever as valorações dos estados de saúde. Numa primeira fase, foram estimados modelos lineares, tanto ao nível individual, como agregado, assim como modelos com efeitos fixos, com efeitos principais e de interacção e com a constante forçada à unidade. Numa segunda fase, foram ainda testados, pelo método dos mínimos quadrados generalizados, modelos com efeitos aleatórios (MEA). Foram também estimados MEA com a constante forçada à unidade pelas equações de estimação generalizadas (EEG). Consideraram-se especificações alternativas, como forma de tomar em consideração o enviesamento das valorações dos estados de saúde. Os modelos foram analisados em termos dos seus coeficientes, ajustamento e capacidade para prever os valores dos estados de saúde captados pelo SG.

Resultados: Os MEA e os modelos EEG produziram coeficientes significativos, que se verificou serem robustos nas especificações dos modelos. A utilização de uma bateria de medidas de bondade do ajustamento levou à conclusão que o modelo EEG com efeitos principais era o melhor modelo estimado.

Conclusão: Esta investigação demonstra que é possível obter sistemas de pesos para a medição da qualidade de vida relacionada com a saúde. O melhor modelo obtido parece prever adequadamente os valores dos estados de saúde da população portuguesa. No entanto, existem algumas incoerências que devem ser investigadas.

Palavras-chave: dados de estados de saúde baseados em preferências; equações de estimação generalizadas; modelos de efeitos aleatórios; qualidade de vida relacionada com a saúde; SF-6D; utilidades dos estados de saúde.

□

Pedro Lopes Ferreira é professor associado na Faculdade de Economia da Universidade de Coimbra e é director do Centro de Estudos e Investigação em Saúde da Universidade de Coimbra. Lara Noronha Ferreira é mestre em Gestão e Economia da Saúde pela Universidade de Coimbra, é professora adjunta na Escola Superior de Gestão, Hotelaria e Turismo da Universidade do Algarve e é membro do Centro de Estudos e Investigação em Saúde da Universidade de Coimbra.

Luís Nobre Pereira é mestre em Estatística e Gestão de Informação pela Universidade Nova de Lisboa e é professor adjunto na Escola Superior de Gestão, Hotelaria e Turismo da Universidade do Algarve.

Entregue em Novembro de 2008

1. Introdução

O SF-6D é um instrumento de medição da qualidade de vida relacionada com a saúde (QVRS) baseado em preferências, recentemente desenvolvido por uma equipa da Universidade de *Sheffield*, que resulta de 11 itens do SF-36 convertidos num sistema de classificação com seis dimensões, cada uma delas com 4 a 6 níveis, permitindo gerar um total de 18 000 estados de saúde diferentes (Brazier, Roberts e Deverill, 2002). O objectivo era permitir utilizar um dos perfis de saúde mais utilizados em todo o mundo, o SF-36, em valorações, isto é, torná-lo num instrumento de medição de preferências. O SF-6D tem seis dimensões: a Função Física, a Limitação no Desempenho, a Função Social, a Dor Física, a Saúde Mental e a Vitalidade. Cada uma das dimensões tem entre 4 a 6 níveis cada, permitindo gerar 18 000 estados de saúde diferentes ($6 \times 4 \times 5 \times 6 \times 5 \times 5$). Uma amostra de 611 indivíduos, representativa da população do Reino Unido, valorizou 249 estados de saúde definidos pelo SF-6D utilizando o *standard gamble* (SG), aplicado por entrevista pessoal (Brazier, Roberts e Deverill, 2002). Esses valores foram utilizados para estimar modelos econométricos que permitem atribuir pesos aos diferentes níveis das dimensões do SF-6D, gerando valores para todos os estados de saúde definidos pelo SF-6D. O índice SF-6D pode ser entendido como um valor contínuo numa escala de 0,30 a 1,00, na qual 1,00 significa «saúde plena» (Brazier, Roberts e Deverill, 2002). Em 2006 foi publicada a versão portuguesa do SF-6D (Ferreira e Ferreira, 2006).

Nos últimos tempos tem-se assistido ao desenvolvimento de estudos em alguns países, de forma a obter funções de valoração relativas aos instrumentos de medição de preferências mais utilizados. No entanto, o número publicado de funções de valoração internacionais mantém-se ainda relativamente pequeno (Brazier *et al.*, 2007). Os primeiros trabalhos publicados utilizando medidas baseadas em preferências tenderam a utilizar os sistemas de valores provenientes dos estudos de valoração do Reino Unido (para o EQ-5D e o SF-6D) e do Canadá para o (*Health Utilities Index* — HUI), dado o elevado custo de levar a cabo estudos de valoração nacionais. No entanto, existe um interesse crescente no estudo da variação entre países das valorações dos estados de saúde e alguma evidência recente sugere que os resultados de um país não têm necessariamente que ser transferidos para outros países (Brazier *et al.*, 2007). Na verdade, as valorações dos estados de saúde podem diferir de país para país devido a diferenças em termos demográficos, valores sócio-culturais e nos sistemas políticos e económicos (Huang *et*

al., 2007; Johnson *et al.*, 1998), existindo ainda quem defenda que sem o uso de pesos apropriados, as inferências sobre os valores de utilidade poderão ser enviesadas (Huang *et al.*, 2007). Teoricamente, os sistemas de pesos deveriam estar preparados para tomar em consideração as diferenças culturais dos instrumentos adaptados, relativamente ao instrumento original (Huang *et al.*, 2007). No entanto, como isso muitas vezes não acontece, a utilização de sistemas de pesos diferentes, como a aplicação do sistema de pesos do Reino Unido e dos Estados Unidos da América no caso do EQ-5D podem gerar *Quality Adjusted Life Years* ganhos bastante diferentes (Johnson *et al.*, 2005).

O instrumento com maior número de sistemas de valores derivados para diferentes países é o EQ-5D, muito devido às acções levadas a cabo pelo grupo EuroQol, existindo sistemas de valores para 13 países europeus (Alemanha, Arménia, Bélgica, Dinamarca, Eslovénia, Espanha, Finlândia, Grécia, Hungria, Noruega, Países Baixos, Reino Unido e Suécia) e cinco países não europeus (Canadá, EUA, Japão, Nova Zelândia e Zimbabué). Recentemente, foram publicados resultados de um projecto desenvolvido com o objectivo de estudar as diferenças nas valorações dos estados de saúde EQ-5D em seis países europeus (Finlândia, Alemanha, Países Baixos, Espanha, Suécia e Reino Unido) e propor um sistema de valores comum aos seis países, incorporando os resultados obtidos nos estudos de valoração nacionais (Greiner *et al.*, 2003). Estão neste momento a ser conduzidos estudos de valoração para o SF-6D no Japão, Hong-Kong, Austrália e Brasil; foi recentemente determinado um sistema de pesos para o HUI2 para o Reino Unido (McCabe *et al.*, 2005) e para o HUI3 para França (Le Galès *et al.*, 2002). Estudos com comparações de sistemas de valores do EQ-5D e do SF-36v2 de diferentes países apontaram para a existência de diferentes valorações dos estados de saúde salientando a importância da existência de sistemas de valores específicos para cada país (Badia *et al.*, 2001; Bussbach *et al.*, 2003; Hawthorne *et al.*, 2007; Huang *et al.*, 2007; Johnson *et al.*, 2000, 2005).

Para Hawthorne *et al.* (2007), a justificação para a derivação de sistemas de valores adaptados ao próprio país reside no facto do conceito de saúde ser uma função multidimensional influenciada pelas expectativas de saúde dos indivíduos e pelo seu próprio estado de saúde e entendida de forma diferente pelos indivíduos. Neste sentido, comparações directas entre países poderão gerar diferenças que se poderão dever a diferenças culturais e não a reais diferenças de saúde. Hawthorne *et al.* (2007) defendem até a existência de diferentes sistemas de pesos

em regiões ou comunidades dentro do mesmo país, se as diferenças culturais existentes entre as regiões ou comunidades forem realmente grandes.

O objectivo deste estudo foi a determinação de um sistema de valores subjacente ao SF-6D para Portugal, através da modelação de dados cardinais obtidos a partir de uma amostra da população portuguesa.

2. Metodologia

2.1. Desenho do estudo

Na valoração dos estados de saúde foi utilizado o SG, considerado o padrão de ouro da medição de utilidades em saúde por se basear nos axiomas da teoria da utilidade esperada (Brazier *et al.*, 2007; Furlong *et al.*, 1990; Torrance, 1986). O SG foi aplicado por entrevista pessoal por entrevistadores treinados, com a duração média de uma hora, tendo os respondentes sido abordados nas suas residências, entre Abril e Outubro de 2006. Com o propósito de obter informação relevante sobre os entrevistados e como exercício introdutório, foi pedida a cada respondente uma descrição da sua própria saúde, através do preenchimento das versões portuguesas do SF-36v2 (Ferreira, 2000a; 2000b), do EQ-5D e do SF-6D (Ferreira e Ferreira, 2006). Foram também colocadas aos indivíduos perguntas com o intuito de recolher dados sociodemográficos, tendo este preâmbulo permitido uma familiarização dos respondentes com o SF-6D. Em seguida, cada respondente ordenou seis estados de saúde definidos pelo SF-6D, acrescidos do melhor (111111) e do pior (645655) estados de saúde possíveis definidos pelo sistema descritivo do SF-6D, para além da morte imediata. Para a valoração dos estados de saúde, foram colocadas aos respondentes seis questões de SG, em que eles tinham que valorizar seis estados de saúde pré-definidos pelo SF-6D num jogo contra o melhor e o pior estado de saúde. Foi também colocada aos indivíduos uma sétima questão de SG, cuja forma dependia de como eles tivessem valorizado o pior estado de saúde, ou seja, como melhor ou pior quando comparado com a morte imediata. Se tivessem considerado o pior estado de saúde definido pelo SF-6D como melhor do que a morte imediata, então ser-lhes-ia apresentada uma escolha entre um resultado incerto de saúde perfeita (111111) ou morte imediata e o resultado certo de ficar no pior estado de saúde (645655). Pelo contrário, se tivessem considerado o pior estado de saúde definido pelo SF-6D como pior do que a morte imediata, então ser-lhes-ia apresentada uma escolha entre um resultado incerto de saúde perfeita ou ficar no pior estado de saúde e o resultado certo de morte imediata. O resul-

tado desta última questão de SG foi utilizado para ligar os estados de saúde de forma a colocá-los na escala de morte-saúde perfeita (0-1) (Brazier Roberts e Deverill, 2002). Esta conversão foi feita com base no trabalho de Patrick *et al.* (1994), tendo-se atribuído o negativo da probabilidade de indiferença do melhor resultado ($-p$) aos estados de saúde valorizados como piores que a morte. Assim, o ajustamento das valorações dos seis estados de saúde intermédios definidos pelo SF-6D foi feito da seguinte forma: $SG_{i,aj} = SG_i + (1 - SG_i)P$, onde $SG_{i,aj}$ é o valor ajustado do estado de saúde i definido pelo SF-6D; SG é o valor do estado de saúde i definido pelo SF-6D; P é o valor do pior estado de saúde definido pelo SF-6D.

2.2. Plano de sondagem

Sendo o objectivo principal deste trabalho a definição de um sistema de valores baseado em preferências para o SF-6D, o universo foi definido como a população portuguesa. Na impossibilidade de estudar toda a população, decidiu-se recolher uma amostra representativa da população portuguesa, em termos de género e idade. No entanto, os elevados custos em que se incorreria para realizar entrevistas pessoais em todo o país fizeram com que se restringisse à população residente na região Algarve (população alvo), mantendo-se a representatividade da população portuguesa em termos de distribuição por género e idade. Por outro lado, considerou-se que os indivíduos com menos de 15 anos poderiam não compreender bem a entrevista, pelo que se decidiu estudar apenas os indivíduos com idade igual ou superior a 15 anos de idade. A definição da população em estudo teve ainda em linha de conta o facto da base de amostragem, passível de estratificação por género e idade, provir de uma empresa de telecomunicações, pelo que teve que se ter em conta que a população em estudo teria que ser obrigatoriamente definida como os residentes na região Algarve, detentores de telefone fixo. Assim, a população inquirida foi definida como os indivíduos residentes na região Algarve, com idade igual ou superior a 15 anos de idade e com telefone fixo. Note-se que não existe nenhuma razão que possa levar a pensar que os resultados seriam muito diferentes se se tivesse utilizado uma amostra da população alargada a todas as regiões de Portugal, principalmente se se tiver em conta que a distribuição da população residente no Algarve é muito semelhante à distribuição da população portuguesa (*Tabela I*).

Foi utilizado um método de sondagem probabilístico que permitisse fazer inferência estatística, quantificar as margens de erro, calcular a dimensão da amostra

adequada à precisão e confiança desejadas para os resultados e tentar evitar os eventuais enviesamentos resultantes do entrevistador. A utilização de informação auxiliar disponível nas publicações do Instituto Nacional de Estatística (Portugal, INE, 2001) sobre a composição da população residente no Algarve em termos de idade e género permitiu dividir a população alvo em 8 estratos mutuamente exclusivos, considerando quatro classes etárias (15-24 anos; 25-49 anos; 50-64 anos; 65 ou mais anos) e o género (feminino e masculino). No que se refere ao processo de selecção das unidades da população que iriam constituir a amostra, decidiu-se utilizar uma amostra estratificada óptima, uma vez que neste tipo de amostragem a taxa de sondagem varia de estrato para estrato, aumentando nos estratos de variância superior, pelo que os estratos com maior variabilidade (i.e. mais heterogéneos) têm maior representação na amostra. Neste método de amostragem a estratégia consiste em contemplar com maior dimensão amostral os estratos que mais contribuem para a variância populacional, i.e., consiste em ir buscar a informação onde ela se encontra (Särndal, Swensson e Wretman, 1992). As opções feitas no plano de sondagem tiveram em conta a simplicidade de implementação, os efeitos sobre a precisão e o custo temporal. A estimativa do desvio padrão em cada estrato da variável de interesse foi obtida a partir da amostra total de um estudo em que se pretendeu obter os valores normais da população activa portuguesa a partir do SF-36v2 (Ferreira e Santana, 2003), tendo-se considerado como variável de interesse a questão 1 do SF-36v2 (em geral, diria que a sua saúde é: Óptima (1); Muito Boa (2); Boa (3); Razoável (4); Fraca (5)). O plano de sondagem previa uma distribuição da amostra pelos estratos, de acordo com a variabilidade exist

tente em cada estrato. Dentro de cada estrato foi utilizada uma sondagem aleatória simples sem reposição. Assim, os indivíduos foram aleatoriamente seleccionados da base de dados que continha dados relativos ao nome, morada, género, idade e número de telefone fixo. A amostra final foi de 140 indivíduos, divididos por 8 estratos (Tabela I).

2.3. Selecção dos estados de saúde

Na impossibilidade de valorizar todos os 18 000 estados de saúde definidos pelo SF-6D, foi necessário seleccionar um número de estados de saúde para serem avaliados. Foram utilizados o mesmo conjunto de estados de saúde avaliados por Brazier e a sua equipa na definição do sistema de valores do Reino Unido (Brazier, Roberts e Deverill, 2002), a fim de possibilitar uma comparação dos resultados desta investigação com aqueles e utilizou-se uma metodologia semelhante à utilizada por aqueles investigadores na selecção dos estados de saúde. O tamanho mínimo da amostra de estados de saúde identificada pelo procedimento do ortoplano do programa estatístico *Statistical Package for the Social Sciences* (SPSS) foi de 49 estados de saúde, pelo que utilizámos os mesmos 49 estados de saúde seleccionados por Brazier. Decidiu-se maximizar o número de valorações por estado de saúde e o número de estados de saúde avaliados, pelo que foram aleatoriamente gerados mais cinco estados de saúde que foram avaliados conjuntamente com os outros 49. Os respondentes avaliaram um total de 54 estados de saúde, para além do pior estado de saúde definido pelo SF-6D, tendo cada estado de saúde sido avaliado por dez indivíduos diferentes. Os estados de

Tabela I
Estratos, população portuguesa, população do Algarve e dimensão amostral

Género	Estratos		População Portuguesa*		População do Algarve*		Amostra
	Idade						
Masculino	15-24 anos		725 770	8%	24 997	7%	11
	25-49 anos		1 939 236	21%	74 565	22%	32
	50-64 anos		889 096	10%	35 768	11%	17
	≥ 65 anos		764 096	8%	33 059	10%	8
Feminino	15-24 anos		702 232	8%	23 761	7%	12
	25-49 anos		1 975 644	22%	71 551	21%	34
	50-64 anos		979 417	11%	35 763	11%	17
	≥ 65 anos		1 054 480	12%	41 085	12%	9
Total	–		9 029 971	100%	340 549	100%	140

* Fonte: Portugal (2001).

saúde foram classificados em suaves, moderados e graves, tendo sido utilizado o método de amostragem estratificado para garantir que a cada indivíduo seria pedida a avaliação de um conjunto de estados de saúde que abarcasse toda a classificação previamente definida. A gravidade dos estados de saúde foi determinada a partir da soma dos níveis das dimensões dos estados de saúde, de acordo com uma abordagem publicada em 2005 (Brazier *et al.*, 2005). Os estados de saúde foram então ordenados de acordo com a soma dos níveis das dimensões e divididos de forma a identificar os três grupos de gravidade. Em seguida, foram aleatoriamente seleccionados, sem reposição, dois estados de saúde de cada grupo, gerando-se desta forma nove conjuntos de estados de saúde. Cada estado de saúde foi avaliado dez vezes, isto é, por dez indivíduos diferentes. Na *Tabela II* é apresentado o conjunto de estados de saúde seleccionados para a tarefa de valoração.

2.4. Análise de dados

Na caracterização da amostra foi realizada uma análise descritiva, tendo-se calculado frequências e medidas de estatística descritiva para as variáveis sociodemográficas e para os valores do SG ajustado. A análise de dados foi realizada no programa estatístico SPSS, versão 15.0.

2.5. Modelação econométrica

Foi seguida a metodologia de Brazier (Brazier, Roberts e Deverill, 2002) no ajustamento de modelos, tanto ao nível individual como agregado, uma vez que se considera que poderá existir um efeito do respondente derivado de variações entre respondentes e dentro dos respondentes. Embora os estados de saúde tenham sido aleatoriamente alocados aos res-

Tabela II
Estados de saúde do SF-6D valorizados

Número de estados de saúde: 54 + pior estado de saúde definido pelo SF-6D

211111	144341
121212	224612
232111	534113
113411	633122
321122	235224
111621	334251
122233	414522
133132	432621
511114	431443
611221	443215
213323	622513
312332	625141
332411	315515
341123	115653
124125	523551
135312	642612
212145	323644
412152	545422
421314	614434
522321	531635
122425	631355
131542	115323
221452	211424
241531	112554
425131	611434
512242	645621
132524	645655
142154	

pondentes, como eles não valorizaram os mesmos estados de saúde, as diferenças entre os valores de estados de saúde poderão estar ligadas às próprias diferenças entre os respondentes e às suas preferências. Ao tomar-se cada indivíduo como uma observação separada poderá ser possível compreender este efeito do respondente. No entanto, como não está provado que modelar ao nível individual é melhor que modelar ao nível agregado, optou-se por seguir os dois caminhos, pelo que se apresentam em seguida os modelos gerais, ao nível individual (1) e ao nível agregado (2).

O modelo geral ao nível individual pode ser definido como:

$$y_{ij} = \alpha + x'_{ij}\beta + r'_{ij}\theta + \varepsilon_{ij} \quad (1)$$

onde $i = 1, 2, \dots, n$ representa os estados de saúde e $j = 1, 2, \dots, m$ representa os respondentes. A variável dependente y_{ij} é o valor ajustado do estado de saúde i valorizado pelo respondente j , $x'_{ij} = (x_{1ij}, x_{2ij}, \dots, x_{vij})$ é um vector de v variáveis *dummy* explicativas referenciadas à mesma unidade, nas quais $x_{vij} = x_{\delta\lambda ij}$ para cada nível λ da dimensão δ do SF-6D. Para qualquer estado de saúde, as variáveis *dummy* explicativas são definidas como 1 se, para este estado de saúde, a dimensão δ está no nível λ , e igual a 0, caso contrário. O nível 1 é a base para cada dimensão. O termo $r'_{ij} = (r_{1ij}, r_{2ij}, \dots, r_{uij})$ é um vector de u variáveis de interacção entre os níveis dos diferentes atributos, também referenciadas à mesma unidade, $\beta' = (\beta_1, \beta_2, \dots, \beta_v)$ e $\theta' = (\theta_1, \theta_2, \dots, \theta_u)$ são vectores de parâmetros e ε_{ij} é uma variável residual. Como o objectivo do estudo assentava na estimação dos valores dos estados de saúde SF-6D para a comunidade, as características pessoais dos respondentes, como a idade, o género, o estado civil e outras, não foram incluídas no modelo.

O modelo geral ao nível agregado pode ser definido como:

$$y_i = \alpha + x'_i\beta + r'_i\theta + \varepsilon_i \quad (2)$$

onde $i = 1, 2, \dots, n$ representa os estados de saúde; y_i é o valor agregado ajustado do estado de saúde i , $x'_i = (x_{1i}, x_{2i}, \dots, x_{vi})$ é um vector de v variáveis *dummy* explicativas referenciadas à mesma unidade, $r'_i = (r_{1i}, r_{2i}, \dots, r_{ui})$ é um vector de u variáveis de interacção entre os níveis dos diferentes atributos, também referenciadas à mesma unidade, β e θ são vectores de parâmetros e ε_i é uma variável residual. Nos modelos gerais, quer ao nível individual (1), quer ao nível agregado (2), poderiam ser introduzidas 300 interacções de primeira ordem entre as variáveis explicativas e milhares de possíveis interacções de ordem

superior. A estimação de todas elas poderia ser possível através de um desenho factorial total ou fraccionário, mas para tal seria necessário uma valoração por parte dos indivíduos de todos os 18 000 estados de saúde gerados pelo SF-6D. Por outro lado, o número elevado de interacções de primeira ordem leva à possibilidade de existência de colinearidade e ao risco de verificar que algumas eram significativas apenas devido ao acaso (Brazier *et al.*, 2007). Tendo em conta as limitações desta investigação, nomeadamente as de carácter financeiro e temporal, optou-se por reduzir o âmbito das interacções introduzidas no modelo seguindo a metodologia aplicada por Brazier (Brazier, Roberts e Deverill, 2002) e por Dolan (Dolan, 1997) em investigações similares. Em ambos os estudos foi incluída no modelo uma variável *dummy* em representação de todas as situações em que qualquer dimensão se encontra no pior nível. Esta variável foi designada por PIOR, por estar ligada aos níveis que denotam situações mais graves e, portanto, piores. A variável *dummy* PIOR foi definida como tomando o valor 1 se qualquer dimensão estiver no nível mais grave e 0 caso contrário.

A interpretação da constante dos modelos de regressão é, no contexto da economia da saúde, algo controversa. De facto, Dolan (1997) num estudo de valoração do EQ-5D atribuiu-lhe uma interpretação estatística baseada no facto de que a constante deveria representar o valor estimado para a saúde perfeita, o que significa que quando todas as *dummies* tomarem o valor zero, então o valor da constante será o valor estimado de 111111. Com efeito, o valor ajustado do SG foi estimado com base nos axiomas da teoria da utilidade esperada, assumindo que o estado 111111 tem o valor 1 e a morte o valor 0. Mais ainda, para que os valores estimados possam ser utilizados em análises custo-utilidade no âmbito da economia da saúde, é necessário admitir que o estado 111111 é equivalente à saúde perfeita e, portanto, tem o valor 1, de acordo com a escala de morte-saúde perfeita (0-1). A melhor forma de garantir isto é forçar a constante a ter um valor unitário, isto é, 1 (Brazier, Roberts e Deverill, 2002). Neste sentido, os modelos ao nível individual e agregado foram também estimados, forçando a constante a ter um valor unitário.

Numa primeira fase, foram estimados modelos lineares, tanto ao nível individual, como agregado. Foram estimados modelos com efeitos fixos, com efeitos principais e de interacção e com a constante forçada à unidade. Detectaram-se, através do teste de Breusch-Pagan, problemas de heteroscedasticidade na estimação de todos os modelos, pelo que todos os modelos foram estimados utilizando o estimador consistente de White.

Numa segunda fase, foram ainda estimados, pelo método dos mínimos quadrados generalizados (MQG), modelos com efeitos aleatórios. O modelo de efeitos aleatórios (MEA), também designado modelo de efeitos mistos ou modelos mistos, em que misto se refere à presença de efeitos fixos e aleatórios, é também hierárquico e multinível, uma vez que as observações se referem a pontos do tempo (nível 1) e se encontram ligadas aos indivíduos (nível 2) (Rabe-Hesketh e Everitt, 2007). Neste sentido, o MEA é apropriado para qualquer tipo de dados agrupados ou binível, não apenas longitudinais (Rabe-Hesketh e Everitt, 2007). Pode, por exemplo, ser adequado em casos em que não se pode assumir que as respostas dos indivíduos do mesmo grupo são independentes porque deve existir heterogeneidade não observada entre os grupos. Neste caso particular, decidiu-se utilizar este modelo por se considerar que as respostas do mesmo indivíduo, relativamente aos diferentes estados de saúde, se encontram agrupadas e não são independentes.

Assim, o MEA pode ser definido como:

$$y_{ij} = \alpha + x'_{ij}\beta + r'_{ij}\theta + u_j + e_{ij} \quad (3)$$

onde $i = 1, 2, \dots, n$ representa os estados de saúde, $j = 1, 2, \dots, m$ representa os respondentes, y_{ij} é o valor ajustado do estado de saúde i valorizado pelo respondente j , $x'_{ij} = (x_{1ij}, x_{2ij}, \dots, x_{vij})$ é um vector de v variáveis *dummy* explicativas referenciadas à mesma unidade, nas quais $x_{vij} = x_{\delta\lambda ij}$ para cada nível λ da dimensão δ do SF-6D. Tal como no modelo (2), para qualquer estado de saúde, as variáveis *dummy* explicativas são definidas como 1 se, para esse estado de saúde, a dimensão δ está no nível λ , e igual a 0, em caso contrário, sendo o nível 1 a base para cada dimensão. O termo $r'_{ij} = (r_{1ij}, r_{2ij}, \dots, r_{vij})$ é um vector de u variáveis inter-acionadas entre os níveis dos diferentes atributos, também referenciadas à mesma unidade, $\beta'_{ij} = (\beta_1, \beta_2, \dots, \beta_v)$ e $\theta'_{ij} = (\theta_1, \theta_2, \dots, \theta_u)$ são vectores de parâmetros. O termo u_j é a variação específica do respondente que se assume variar aleatoriamente entre os respondentes e ε_{ij} é um termo de erro da i -ésima avaliação do estado de saúde pelo indivíduo j , assumindo-se que varia aleatoriamente entre as observações com $e_{ij} \sim [0, \sigma^2]$. Para além disto, $Cov(u_j, e_{ij}) = 0$, o que significa que a alocação dos estados de saúde pelos respondentes é aleatória. Foi utilizado o teste de Hausman para testar a especificação do modelo e comparar o modelo de efeitos fixos com o MEA. Analogamente foi utilizado o teste Breush-Pagan dos multiplicadores de Lagrange (B-P LM) para comparar o modelo estimado pelo método dos mínimos quadrados ordinários (MMQ) com o MEA.

Numa terceira fase, foram estimados MEA com a constante forçada à unidade pelas equações de estimação generalizadas (EEG), que é uma abordagem largamente utilizada na análise de dados longitudinais em estudos clínicos e epidemiológicos. Foi proposta por Liang e Zeger (1986) e Zeger e Liang (1986) e é uma extensão do modelo linear generalizado para dados correlacionados. Os modelos EEG permitem a especificação da estrutura de correlação dentro do grupo para os painéis e são também conhecidos como modelos de dados em painel da média da população (*population-averaged panel-data model*). Todos os modelos EEG consideram uma equação de estimação escrita em duas partes. A primeira parte estima os parâmetros de regressão e a segunda estima a associação entre os parâmetros de associação da distribuição de variância de segunda ordem (Hardin e Hilbe, 2003). Tal como nos modelos lineares generalizados, é necessário especificar uma distribuição para a componente aleatória (que determina a função de variância) e uma função de ligação (entre a variável e o seu regressor linear) para transformar os valores esperados. Estes modelos EEG permitem observações repetidas ou dependentes e têm a capacidade de especificar a estrutura de correlação. Nos modelos EEG os parâmetros são estimados utilizando as «equações de estimação», que se assemelham às equações *score* da estimação pela máxima verosimilhança do MEA. Estas equações de estimação apenas requerem especificação de uma ligação, de uma função de variância e de uma estrutura de correlação para as respostas observadas condicionadas às covariáveis (Rabe-Hesketh e Everitt, 2007). Para a estimação dos MEA com a constante forçada à unidade, utilizaram-se modelos EEG, considerando-se a distribuição gaussiana, a identidade como função de ligação e uma estrutura de correlação permutável. O teste de especificação do modelo e a comparação entre o modelo de efeitos fixos e o modelo EEG foi feita através do teste de Hausman.

Como a distribuição da variável dependente era enviesada, decidiu-se considerar especificações alternativas, com o objectivo de melhorar o desempenho preditivo do modelo. Neste sentido, fizeram-se três transformações da variável dependente: uma transformação *logit* e duas transformações *log-log*. A transformação *log-log* apresentada neste artigo foi a seguinte: $\ln[-\ln(1 - 1/2(SG_{aj} + 1))]$, onde SG_{aj} é o valor SG ajustado à escala 0-1 (morte-saúde perfeita), também usada nos outros modelos. As transformações foram modeladas com efeitos aleatórios utilizando o MQG. Mais uma vez, foram utilizados o teste de Hausman e o teste B-P LM para comparar, respectivamente, o modelo de efeitos fixos e o modelo estimado pelo MMQ com o MEA.

Nos resultados dos modelos, espera-se que os coeficientes estimados sejam negativos e que aumentem em grandeza absoluta, uma vez que as variáveis *dummy* representam problemas progressivamente piores em cada dimensão, quando comparadas com a base de cada dimensão (Brazier, Roberts e Deverill, 2002). Pela mesma razão, considera-se que ocorre uma incoerência quando um coeficiente estimado diminui em valor absoluto, quando se passa para um nível mais grave de uma dada dimensão. Neste sentido, um bom modelo deverá maximizar o número de coeficientes negativos e minimizar o número de incoerências.

O desempenho preditivo dos modelos foi analisado utilizando a seguinte bateria de medidas: o coeficiente de determinação ajustado (R_{aj}^2), o erro absoluto médio (EAM) e a percentagem de observações correctamente previstas a um erro absoluto inferior a 0,1 ($\%EA < 10,10$) e a 0,05 ($\%EA < 10,05$). As medidas de bondade do ajustamento utilizadas incluíram o critério de informação de Akaike (*Akaike's information criterion*) (AIC) e o critério de informação de Bayes (*Bayes information criterion*) (BIC). No caso das EEG, que se baseiam na teoria da quase-verosimilhança, as medidas AIC e BIC não são directamente aplicáveis (Cui e Qian, 2007), pelo que se utilizou uma alternativa ao critério AIC para dados correlacionados proposta por Pan (2001), o *quasilikelihood under the independence model information criterion* (QIC), que pode ser utilizado nos modelos EEG como medida da qualidade do ajustamento e critério

de ajuda na selecção do melhor modelo (Craiu, Duchesne e Fortin, 2008; Cui e Qian, 2007). Foram ainda testadas a nulidade da média dos resíduos [$t(\text{média} = 0)$] e a incorrecta especificação do modelo (teste RESET de Ramsey — *Ramsey's RESET Test*). A normalidade dos resíduos foi testada através do teste assintótico de Jarque-Bera (JB), do teste de assimetria e curtose para a normalidade (*Skewness and Kurtosis tests for Normality*) e da representação gráfica da distribuição dos resíduos, usando a estimação de uma função de densidade de Kernel. Foi também testada a significância global dos modelos, utilizando um teste de Wald ($W(\chi^2)$) para os MEA e o teste F para todos os outros modelos. A modelação econométrica foi feita no programa Stata, versão 9.0.

3. Resultados

3.1. Caracterização da amostra

Na *Tabela III* são apresentadas as principais características da amostra.

A amostra foi maioritariamente constituída por indivíduos do sexo feminino. Relativamente à idade, a média de idades dos indivíduos era de 42,7 anos (DP = 16,7 anos), sendo que o indivíduo mais novo tinha 15 anos e o mais velho 87. Os homens eram em média mais novos ($\bar{X} = 41,7$;DP = 17,4) do que as mulheres ($\bar{X} = 43,5$;DP = 16,4). Distribuindo os indi-

Tabela III
Características da amostra

Amostra (n)	140	
Mulheres (%)	75	(53,6)
Idade média (DP)	43	(16,9)
Idade entre 25 e 49 anos (%)	66	(47,1)
Casados/união de facto (%)	86	(61,4)
Ensino superior (%)	54	(38,6)
Quadros superiores, dirigentes e especialistas (%)	56	(40,6)
Residentes em zonas urbanas (%)	122	(87,1)
Rendimento entre 1000 euros e 1999 euros (%)	54	(39,4)
Sem doença crónica (%)	90	(64,3)
EQ-5D*: média (DP)	0,86	(0,17)
Escala visual analógica do EQ-5D*: média (DP)	0,80	(0,14)

* Valores calculados a partir do algoritmo do EQ-5D (Dolan, 1997).

DP — Desvio-padrão.

víduos por classes etárias, 47,1% pertenciam à classe entre os 25 e os 49 anos (classe modal), 24,3% dos indivíduos tinham entre 50 e 64 anos e 12,1% pelo menos 65 anos. Dos indivíduos que constituíam a amostra, a maioria era casada ou vivia em união de facto (61,4%). Relativamente às habilitações literárias, os indivíduos com ensino superior representavam 38,6% (54), embora 17,1% (24) possuísem habilitações literárias ao nível do ensino primário. A partir da *Tabela III* é possível observar, na amostra inquirida, a existência de 31,2% de inquiridos quadros superiores, dirigentes e especialistas das profissões intelectuais e científicas e de uma maioria de inquiridos (64,3%) sem doença crónica. No que diz respeito ao local de residência, a maioria residia em zonas urbanas e quanto ao rendimento médio mensal líquido, quase 40% dos indivíduos auferiam um rendimento entre 1000 euros e 1999 euros. O estado de saúde da amostra, utilizando o índice EQ-5D, era de

0,86, enquanto que utilizando a escala visual analógica do EQ-5D esse valor médio era de 0,80, o que denota um bom estado de saúde.

3.2. Valores dos estados de saúde

Após a aplicação do algoritmo de cálculo do SG (Furlong *et al.*, 1990), foram calculados os valores atribuídos por cada indivíduo a cada estado de saúde. Na *Tabela IV* são apresentadas as medidas descritivas dos valores das utilidades dos 55 estados de saúde avaliados pelos indivíduos.

A média da avaliação do conjunto dos estados de saúde foi de 0,595 (DP = 0,290). Em termos globais, os estados de saúde foram valorizados entre -0,500 (645655) e 0,993 (121212; 412152; 213323; 212145) e os valores médios dos valores atribuídos aos estados de saúde variaram entre 0,217 (614434) e 0,831

Tabela IV
Medidas descritivas dos valores SG dos estados de saúde do SF-6D

Estado	\bar{X}	DP	Me	Min	Max	Estado	\bar{X}	DP	Me	Min	Max
111621	0,639	0,203	0,639	0,325	0,958	332411	0,784	0,134	0,813	0,533	0,938
112554	0,737	0,218	0,816	0,278	0,978	334251	0,576	0,270	0,618	0,058	0,948
113411	0,785	0,099	0,820	0,618	0,878	341123	0,725	0,157	0,698	0,483	0,938
115323	0,778	0,222	0,868	0,278	0,978	412152	0,758	0,156	0,730	0,483	0,993
115653	0,608	0,161	0,665	0,325	0,798	414522	0,364	0,244	0,336	-0,013	0,675
121212	0,771	0,152	0,780	0,483	0,993	421314	0,788	0,131	0,808	0,588	0,950
122233	0,528	0,198	0,426	0,325	0,798	425131	0,575	0,166	0,571	0,325	0,808
122425	0,746	0,223	0,835	0,278	0,988	431443	0,589	0,209	0,618	0,278	0,808
124125	0,625	0,182	0,624	0,363	0,863	432621	0,698	0,265	0,793	0,023	0,938
131542	0,542	0,266	0,448	0,123	0,963	443215	0,798	0,120	0,808	0,663	0,948
132524	0,725	0,136	0,683	0,528	0,903	511114	0,666	0,235	0,618	0,348	0,958
133132	0,539	0,193	0,494	0,325	0,798	512242	0,711	0,261	0,818	0,138	0,938
135312	0,668	0,201	0,681	0,325	0,958	522321	0,629	0,182	0,705	0,313	0,825
142154	0,560	0,252	0,448	0,258	0,963	523551	0,733	0,132	0,698	0,588	0,913
144341	0,762	0,107	0,763	0,618	0,903	531635	0,544	0,253	0,521	0,188	0,983
211111	0,715	0,124	0,731	0,533	0,863	534113	0,487	0,168	0,426	0,275	0,730
211424	0,812	0,208	0,865	0,278	0,988	545422	0,596	0,290	0,683	0,123	0,878
212145	0,693	0,257	0,728	0,253	0,993	611221	0,726	0,133	0,763	0,533	0,918
213323	0,734	0,196	0,780	0,368	0,993	611434	0,774	0,227	0,853	0,278	0,978
221452	0,660	0,262	0,785	0,138	0,913	614434	0,217	0,314	0,048	-0,088	0,640
224612	0,750	0,142	0,753	0,533	0,938	622513	0,593	0,306	0,615	0,023	0,978
232111	0,831	0,120	0,828	0,663	0,968	625141	0,500	0,293	0,581	-0,063	0,843
235224	0,620	0,169	0,609	0,368	0,863	631355	0,658	0,289	0,689	0,023	0,978
241531	0,496	0,231	0,426	0,130	0,798	633122	0,451	0,247	0,533	0,020	0,843
312332	0,746	0,126	0,763	0,528	0,878	642612	0,408	0,322	0,363	-0,148	0,873
315515	0,710	0,179	0,725	0,363	0,968	645621	0,674	0,269	0,713	0,278	0,978
321122	0,631	0,279	0,618	0,123	0,958	645655	0,278	0,363	0,300	-0,500	0,950
323644	0,677	0,278	0,781	0,080	0,913						

X — Média; DP — Desvio Padrão; Me — Mediana; Min — Mínimo; Max — Máximo.

(232111). A mediana dos valores atribuídos aos estados de saúde variou entre 0,048 (614434) e 0,868 (115323). Nenhum indivíduo atribuiu o valor 1,0 a um estado de saúde, embora tenham sido feitas 120 (19,0%) avaliações superiores a 0,85. De facto, no que concerne às avaliações efectuadas pelos indivíduos, verificaram-se cerca de 4,0% de avaliações negativas e cerca de 7,0% de avaliações entre 0,00 e 0,25. É de referir que mais de 50,0% das avaliações foram feitas no topo da escala, isto é, atribuindo aos estados de saúde valores superiores a 0,65.

Como é possível observar a partir da *Tabela V*, alguns indivíduos atribuíram valores negativos a

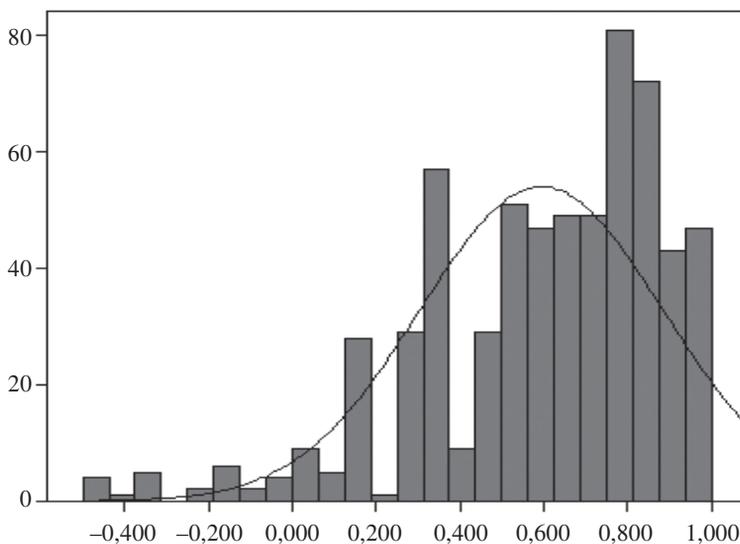
alguns estados de saúde: 17,1% (24) dos indivíduos atribuíram valores negativos, que variaram entre -0,500 e -0,013, a 10,9% (6) dos estados de saúde avaliados. A identificação dos estados a que os indivíduos atribuem valores negativos é importante, uma vez que denota quais os estados de saúde considerados piores que a morte.

A partir da *Tabela IV* é, ainda, possível observar a assimetria da distribuição, uma vez que os valores da mediana excedem os da média na maior parte dos casos. Esta assimetria é ainda mais evidente ao nível individual. De facto, a distribuição é fortemente assimétrica negativa, como se pode observar pela *Figura 1*

Tabela V
Estados de saúde avaliados com valores negativos e número de indivíduos que os atribuíram

Estado de saúde	SG ajustado	Número de indivíduos
645655	-0,500	2
645655	-0,450	2
645655	-0,400	1
645655	-0,350	5
645655	-0,250	1
645655	-0,200	1
645655	-0,150	5
642612	-0,148	1
614434	-0,088	1
614434	-0,080	1
625141	-0,063	1
614434	-0,050	1
614434	-0,015	1
414522	-0,013	1

Figura 1
Histograma da distribuição dos valores ajustados dos estados de saúde (SG ajustado)



($g = -1,09$). No que respeita à curtose, a distribuição é leptocúrtica, conforme também se pode concluir pela observação do histograma e pelo valor do coeficiente de achatamento ($g_2 = -1,16$).

Estes resultados indicam que existe um grande número de indivíduos com valores elevados no SG, concentrados num intervalo de pequena amplitude. A distribuição apresenta quatro *outliers* moderados, que representam indivíduos que atribuíram valorações

bastante abaixo dos demais (valores menores ou iguais a $-0,350$).

3.3. Modelação dos valores dos estados de saúde

Foram estimados vários modelos, mas apenas se apresentam nas *Tabelas VI e VII* os melhores modelos, em termos de especificação e de interpretação.

Tabela VI
Modelos lineares

	Nível agregado (média)		Nível Individual	
	M1 (efeitos principais)	M2 (efeitos de interação)	M3 (efeitos principais)	M4 (efeitos de interação)
<i>c</i>	1,000 ^d	1,000 ^d	1,000 ^d	1,000 ^d
FF2	-0,019	-0,026	0,003	0,000
FF 3	0,011	0,009	0,034 ^b	0,034 ^b
FF 4	-0,022	-0,027	-0,021	-0,020
FF 5	-0,047	-0,046	-0,061 ^b	-0,058 ^b
FF 6	-0,174 ^c	-0,189 ^c	-0,190 ^d	-0,190 ^d
LD2	-0,022	-0,031	-0,016	-0,021
LD 3	-0,079 ^b	-0,080 ^b	-0,064 ^d	-0,065 ^d
LD 4	-0,070	-0,085	-0,101 ^d	-0,109 ^d
FS2	0,024	0,018	0,003	0,005
FS 3	0,004	-0,001	0,008	0,010
FS 4	-0,168 ^c	-0,177 ^c	-0,164 ^d	-0,165 ^d
FS 5	-0,069 ^b	-0,088 ^b	-0,090 ^d	-0,096 ^d
DF2	0,079 ^b	0,077 ^b	0,070 ^c	0,068 ^c
DF 3	0,126 ^c	0,125 ^c	0,133 ^d	0,140 ^d
DF 4	0,053	0,053	0,052 ^c	0,055 ^c
DF 5	-0,017	-0,023	-0,031	-0,027
DF 6	-0,008	-0,021	-0,034	-0,033
SM2	-0,030	-0,041	-0,042 ^c	-0,047 ^c
SM 3	-0,161 ^d	-0,166 ^d	-0,185 ^d	-0,189 ^d
SM 4	-0,084 ^b	-0,090 ^b	-0,054 ^c	-0,062 ^c
SM 5	-0,065	-0,085 ^b	-0,095 ^d	-0,104 ^d
VT2	-0,091 ^b	-0,087 ^b	-0,078 ^d	-0,076 ^d
VT3	-0,082 ^b	-0,075 ^b	-0,074 ^d	-0,074 ^d
VT4	-0,046	-0,046	-0,005	-0,010
VT5	-0,014	-0,029	-0,059 ^d	-0,067 ^d
PIOR	-	0,039	-	0,025
<i>n</i>	55	55	630	630
Incoerências	7	6	7	7
EAM	0,061	0,059	0,197	0,196
%EA < 10,051	65,45	63,64	15,87	15,87
%EA < 10,101	78,18	78,18	29,84	29,68
<i>t</i> (mean = 0)	- ^a	- ^a	-0,597	-0,575
RESET	5,36 ^c	4,84 ^c	5,24 ^c	2,77 ^b
F	287,14 ^d	273,81 ^d	1369,53 ^d	1322,29 ^d
JB	2,954	2,926	36,39 ^d	36,17 ^d
AIC	-146,573	-145,960	-1223,112	-1223,776
BIC	-94,383	-91,762	-1107,524	-1103,741

^a O erro médio é zero por definição; ^b $p < 0,10$; ^c $p < 0,01$; ^d $p < 0,001$.

FF — Função física; LD — Limitação de desempenho; FS — Função social; DF — Dor física; SM — Saúde mental; VT — Vitalidade.

Nos modelos média ao nível agregado (M1 e M2), a maior parte dos coeficientes têm o sinal esperado negativo. Dos 25 coeficientes do M1, 11 são significativos e existem 7 incoerências. A inclusão do termo de interação não melhorou significativamente o modelo, embora tenha reduzido o número de incoerências e aumentado o número de coeficientes significativos (12). Quando comparado com o modelo semelhante relatado em Brazier, Roberts e Deverill (2002), o M1 apresenta menos coeficientes significativos (11 *versus* 24), mais incoerências (7 *versus* 5) e mais coeficientes significativos positivos (2 *versus* 0).

rências e aumentado o número de coeficientes significativos (12). Quando comparado com o modelo semelhante relatado em Brazier, Roberts e Deverill (2002), o M1 apresenta menos coeficientes significativos (11 *versus* 24), mais incoerências (7 *versus* 5) e mais coeficientes significativos positivos (2 *versus* 0).

Tabela VII
Modelos MEA e EEG (nível individual: n=630)

	MEA		EEG		Transformação Log-Log (MEA)
	M5 (efeitos principais)	M6 (efeitos de interação)	M7 (efeitos principais)	M8 (efeitos de interação)	M9 (efeitos principais)
<i>c</i>	0,827 ^c	0,817 ^c	1,000 ^c	1,000 ^c	1,000 ^c
FF2	-0,041 ^a	-0,040 ^a	-0,050 ^b	-0,050 ^a	-0,065 ^a
FF3	-0,025	-0,024	-0,032	-0,031	-0,041
FF4	-0,040 ^a	-0,036 ^a	-0,049 ^a	-0,045 ^a	-0,066 ^a
FF5	-0,045 ^a	-0,042 ^a	-0,055 ^a	-0,052 ^a	-0,069 ^a
FF6	-0,177 ^c	-0,179 ^c	-0,214 ^c	-0,220 ^c	-0,320 ^c
LD2	-0,026	-0,028 ^a	-0,031	-0,034 ^a	-0,065 ^a
LD3	-0,003	-0,003	-0,003	-0,003	-0,014
LD 4	-0,046 ^a	-0,056 ^b	-0,054 ^a	-0,067 ^b	-0,112 ^b
SM2	-0,031 ^a	-0,033 ^a	-0,038 ^a	-0,042 ^a	-0,055 ^a
SM3	-0,012	-0,015	-0,014	-0,019	-0,032
SM4	-0,034	-0,052 ^a	-0,039	-0,060 ^a	-0,095 ^a
SM5	-0,057 ^b	-0,066 ^b	-0,069 ^b	-0,080 ^b	-0,114 ^b
DF2	0,006	0,007	0,006	0,008	0,011
DF3	0,001	0,007	-0,001	0,006	0,127
DF4	-0,049 ^a	-0,047 ^a	-0,061 ^a	-0,060 ^a	-0,074 ^a
DF5	-0,044 ^a	-0,050 ^a	-0,054 ^a	-0,060 ^a	-0,096 ^b
DF6	-0,073 ^c	-0,073 ^c	-0,090 ^c	-0,090 ^c	-0,127 ^c
SM2	-0,048 ^b	-0,054 ^b	-0,059 ^b	-0,066 ^b	-0,095 ^b
SM3	-0,011	-0,025	-0,009	-0,026	-0,045
SM4	-0,057 ^b	-0,059 ^b	-0,070 ^b	-0,073 ^b	-0,101 ^b
SM5	-0,085 ^c	-0,092 ^c	-0,103 ^c	-0,112 ^c	-0,180 ^c
VT2	-0,043 ^c	-0,040 ^a	-0,051 ^b	-0,048 ^b	-0,070 ^a
VT3	-0,031	-0,026	-0,036	-0,031	-0,059 ^a
VT4	-0,037 ^b	-0,039 ^b	-0,046 ^b	-0,048 ^b	-0,070 ^b
VT5	-0,080 ^c	-0,082 ^c	-0,097 ^c	-0,101 ^c	-0,136 ^c
PIOR	-	0,033 ^a	-	0,038 ^a	0,054 ^a
Ajustado	0,194	0,208	-	-	-
Incoerências	5	5	5	5	3
EAM	0,209	0,207	0,209	0,207	0,225
%EA < 10,051	14,44	14,13	14,29	14,29	13,02
%EA < 10,011	28,25	28,57	28,41	28,25	27,94
<i>t</i> (mean = 0)	0,000	0,000	0,000	0,000	5,419 ^c
RESET	14,39 ^c	7,44 ^b	14,65 ^c	14,46 ^c	15,47 ^c
W(χ^2)	703,05 ^c	711,29 ^c	762,57 ^c	769,61 ^c	786,38 ^c
JB	25,72 ^c	25,37 ^c	25,63 ^c	25,28 ^c	0,19
QIC	-	-	68,069	68,231	153,387

^a $p < 0,10$; ^b $p < 0,01$; ^c $p < 0,001$.

FF — Função física; LD — Limitação no desempenho; FS — Função social; DF — Dor física; SM — Saúde mental; VT — Vitalidade.

Em termos de capacidade preditiva, o M1 tem 65,45% de estados de saúde com previsões com um erro absoluto médio inferior a 5%, enquanto que o modelo de Brazier só tem 52,61%. A comparação do M2 com o correspondente modelo de Brazier, que foi considerado o melhor modelo dos estimados por aqueles autores e que serviu de base à definição do sistema de valores do Reino Unido, indica que, em termos de incoerências, são muito semelhantes (6 *versus* 5), mas em termos de coeficientes significativos com sinal positivo (2 *versus* 0) e de coeficientes significativos (12 *versus* 24) o modelo de Brazier apresenta melhores resultados do que o modelo 8. O M2 tem 63,64% de estados de saúde com previsões com um erro absoluto médio inferior a 5% e 78,18% inferior a 10%, enquanto que o modelo de Brazier tem 51,81% e 79,52%, respectivamente.

O modelo linear ao nível individual com efeitos de interação (M4) tem 18 coeficientes negativos e 18 coeficientes significativos, embora tenha 7 incoerências. Comparando os modelos com base na sua capacidade de predição, verifica-se que o M3 e o M4 têm um comportamento pior, dado que o M1 e M2 têm um EAM menor e mais observações correctamente previstas. Brazier, Roberts e Deverill (2002) não disponibilizam os frutos de modelos semelhantes, pelo que não é possível estabelecer uma comparação. O teste Reset evidencia problemas de especificação. As previsões dos M3 e M4 são não enviesadas e os erros de predição do M1 e do M2 estão normalmente distribuídos (ao nível de significância de 5%). Comparando os quatro modelos em termos de AIC e BIC, verifica-se que os M3 e M4 apresentam um melhor resultado.

Os MEA estimados pelos MQG produziram melhores resultados do que os modelos estimados pelos MMQ. De facto, M5 e M6 têm 23 coeficientes com o sinal negativo esperado e apenas 5 incoerências. Mais ainda, a junção de efeitos de interação ao MEA (M6) aumentou o número de coeficientes significativos de 17 para 20. Estabelecendo uma comparação com os resultados do M5 com um modelo igual apresentado em Brazier *et al.* (2002), verifica-se que apresenta o mesmo número de coeficientes significativos positivos que o de Brazier (0), o mesmo número de coeficientes significativos e sensivelmente o mesmo poder explicativo ($R^2_{aj} = 0,194$) contra ($R^2_{aj \text{ Brazier}} = 0,200$). No entanto, o seu EAM é superior ao do modelo de Brazier ($EAM_{\text{Brazier}} = 0,073$) e apresenta uma capacidade de predição pior ($\%EA < 10,05|_{\text{Brazier}} = 96,50\%$; $\%EA < 10,10|_{\text{Brazier}} = 98,49\%$). O M6 não pode ser comparado com os resultados apresentados em Brazier, Roberts e Deverill (2002), por aqueles autores não apresentarem um modelo

semelhante. Em ambos os modelos, o teste RESET é significativo, indicando uma incorrecta especificação dos modelos e para ambos os modelos os coeficientes considerados no seu conjunto são significativos. Quanto aos resíduos, os resultados do teste JB indicam que não seguem uma distribuição normal, embora as previsões não sejam enviesadas. Note-se que em ambos os modelos a maior parte da variância não explicada é atribuível ao nível individual, indicando um elevado grau de heterogeneidade individual. Esta evidência de heterogeneidade fornece uma justificação forte para a utilização de uma abordagem com modelos de dados em painel (Jones, 2007; Jones *et al.*, 2007). Os resultados do teste de Hausman apontaram para a utilização do MEA, em detrimento do modelo de efeitos fixos, tanto nos casos do M5 como do M6. O teste B-P LM indicou igualmente a utilização do MEA em vez de uma especificação pelo MMQ.

A partir da *Tabela VII*, verifica-se que, no M7, todos os coeficientes significativos têm o sinal esperado negativo, a maioria é significativa e que este modelo tem apenas 5 incoerências. Estes resultados evidenciam um modelo bastante bom, uma vez que apenas as incoerências poderão ser consideradas uma limitação do modelo. Contudo, estas incoerências não constituem um problema real, uma vez que se considera que, à semelhança do que é afirmado por Brazier, Roberts e Deverill (2002), não terão sérias implicações no desempenho do modelo estimado, excepto uma redução da sensibilidade nos extremos superiores das dimensões Função Física, Função Social, Saúde Mental e Vitalidade e nos níveis intermédios da dimensão Dor Física. Comparando com o seu correspondente modelo em Brazier, Roberts e Deverill (2002), ambos apresentam a totalidade dos coeficientes negativos. O M7 só é ligeiramente superado no número de incoerências (5 contra 4) e no número de coeficientes significativos (17 contra 26). No entanto, no que diz respeito à capacidade preditiva o modelo de Brazier é superior ao modelo 15 ($\%EA < 10,05|_{\text{Brazier}} = 96,53\%$; $\%EA < 10,10|_{\text{Brazier}} = 98,32\%$), tendo também melhores resultados em termos de EAM ($EAM_{\text{Brazier}} = 0,078$).

A introdução de efeitos de interação no M7, forçando a constante a ter um valor unitário, permitiu a estimação do M8, embora não tenha melhorado significativamente os resultados: apenas levou a um aumento do número de coeficientes significativos de 17 para 20. Estes resultados evidenciam também um modelo bastante bom, uma vez que apenas as incoerências poderão ser consideradas uma limitação do modelo. Estabelecendo uma analogia com o seu correspondente modelo em Brazier, Roberts e Deverill (2002), constata-se que os resultados são, uma vez

mais, iguais em termos de coeficientes significativos negativos (todos) e do número de coeficientes significativos (20 contra 25). Embora, o M8 tenha menos uma incoerência que o de Brazier, apresenta piores resultados em termos de capacidade preditiva ($\%EA < 10,05|_{Brazier} = 96,66\%$; $\%EA < 10,10|_{Brazier} = 98,55\%$); e um EAM superior ($EAM_{Brazier} = 0,076$). Ambos os modelos apresentaram uma aderência global significativa, embora tenham problemas de especificação. Embora os resíduos sejam não enviesados, não seguem uma distribuição normal. Uma análise dos valores do QIC indica a escolha do M7, em relação ao M8.

O M9 apresenta resultados similares ao M7 no que diz respeito ao número de coeficientes com o sinal esperado negativo, menos incoerências e mais coeficientes significativos. Contudo, o QIC do M7 é menor que o do M9. As previsões do M9 são enviesadas, embora os seus resíduos sigam uma distribuição normal. Os resultados do teste RESET apontam para uma má especificação do M9, embora os do teste de Wald indiquem que os coeficientes, quando tomados no seu conjunto, são significativos. Não é possível estabelecer comparações entre o M9 e os resultados de Brazier (Brazier, Roberts e Deverill, 2002), porque aqueles autores não apresentam os resultados a que chegaram. Apenas referem que estas especificações alternativas não melhoraram significativamente os resultados dos modelos (Brazier, Roberts e Deverill, 2002), o que de facto também se verificou neste estudo.

4. Conclusão

Em termos globais, existem duas abordagens no que diz respeito à equivalência dos sistemas de pesos dos instrumentos de medição da QVRS (Hawthorne *et al.*, 2007). Por um lado, é defendida uma tradução rigorosa e cuidada, seguindo os trâmites correctos, que assegura uma equivalência dos instrumentos entre culturas e, portanto, dispensa a derivação de sistemas de pesos para cada país. Esta é a abordagem defendida pelos autores do SF-36 e pelo grupo do projecto *International Quality of Life Assessment* (IQOLA) (Bullinger *et al.*, 1998; Gandek e Ware, 1998), embora apresente claras dificuldades, na medida em que o sistema descritivo e de valores podem estar culturalmente ligados a uma determinada sociedade e serem, portanto, inapropriados noutras (por exemplo, onde andar é o meio normal de transporte). Por outro lado, outros autores defendem que para um sistema descritivo ser válido entre culturas diferentes é necessário que este seja desenvolvido internacionalmente, mas valorizado com um sis-

tema de pesos local. Esta é a abordagem defendida pela Organização Mundial de Saúde e subjacente aos seus instrumentos (Szabo *et al.*, 1997; The WHOQOL Group, 1995).

É neste contexto que este estudo foi desenvolvido, tendo como objectivo a determinação de um sistema de valores subjacente ao SF-6D para Portugal, através da modelação de dados cardinais obtidos a partir de uma amostra da população portuguesa.

Os resultados alcançados proporcionam um sistema de valores adaptado às especificidades e cultura da população portuguesa e, portanto, uma alternativa ao sistema de valores britânico do SF-6D. Os MEA, tanto com efeitos de interacção (M6), como os estimados pelas EEG, com efeitos principais e com a constante forçada à unidade (M7) foram os considerados melhores após a conjugação dos itens: número de coeficientes com sinal negativo esperado e de coeficientes significativos, número de incoerências, poder explicativo e EAM, especialmente se forem comparados com o desempenho de outras medidas de medição da QVRS (Dolan, 1997; McCabe *et al.*, 2005). No que diz respeito à capacidade de interpretação, o M7 foi considerado o de mais fácil interpretação por duas razões principais. Primeiro, porque tem todos os coeficientes significativos negativos (e mesmo dos não significativos, apenas o segundo nível da Dor Física não é negativo), o que o torna aplicável em análises custo-utilidade. Segundo, porque as incoerências que possui não constituem um problema sério, uma vez que apenas reduzem a sensibilidade nos limites superiores da Função Física, da Função Social, da Saúde Mental e da Vitalidade e nos níveis intermédios da Dor Física. Na realidade, estas incoerências poderão ter a sua razão de ser no facto dos indivíduos não fazerem correctamente a distinção entre, por exemplo, os níveis 2 e 3 da Função Social, tal como foi alvitado por Brazier, Roberts e Deverill (2002) para o sistema de valores britânico. Não obstante, é necessária mais investigação a fim de conhecer as verdadeiras implicações destas incoerências. A sugestão de Brazier, Roberts e Deverill. (2002) para a solução dos seus problemas relativos às cinco incoerências do modelo britânico é válida para este: um aumento da amostra e do número de estados de saúde avaliados poderia levar à resolução do problema.

No entanto, embora o M7 seja considerado um bom modelo, este estudo levanta algumas preocupações de natureza metodológica que merecem continuar a ser analisadas. Primeiro, a amostra de estados de saúde valorizados ($n = 55$) é pequena e representa apenas 0,31% do total de estados de saúde gerados pelo SF-6D, embora na valoração do EQ-5D tenham apenas sido valorizados 43 estados de saúde (Szende

et al., 2007) e Brazier, Roberts e Deverill (2002) considere que o procedimento do ortoplano requer apenas 49 estados de saúde (do total de 18.000) para estimar um modelo aditivo. Segundo, a amostra aleatória de indivíduos que valorizaram os estados de saúde poderá também ser considerada pequena, uma vez que o sistema de valores subjacente ao EQ-5D para o Reino Unido foi estimado a partir de uma amostra com 2.997 indivíduos que valorizaram 17,3% dos possíveis estados de saúde e na estimação do sistema de valores do SF-6D para o mesmo país utilizou-se uma amostra de 611 indivíduos que valorizaram 1,4% dos estados de saúde gerados pelo instrumento (McCabe *et al.*, 2005). Terceiro, a amostra utilizada residia na região Algarve, pelo que os resultados poderão, eventualmente, traduzir este facto. A utilização de uma amostra maior e a valorização de um número mais elevado de estados de saúde poderiam, certamente, contribuir para um modelo com uma capacidade preditiva melhor. No entanto, deverá ter-se em conta que: (i) ambos os estudos referidos anteriormente pretendiam estimar modelos para a população do Reino Unido, bastante maior que a de Portugal; (ii) o EQ-5D apenas gera 243 estados de saúde, razão pela qual a valorização de somente 43 estados de saúde (menos do que os valorizados neste trabalho) corresponde a uma percentagem relativamente elevada do total de estados de saúde; (iii) não existe nenhuma razão que possa levar a pensar que os resultados seriam diferentes se se tivesse utilizado uma amostra da população alargada a todas as regiões de Portugal, principalmente se se tiver em conta que a distribuição da população residente no Algarve é semelhante à distribuição da população portuguesa; (iv) o primeiro trabalho de Brazier no sentido de estimar um sistema de valores para o SF-6D utilizou uma amostra por conveniência de 165 profissionais da saúde da Faculdade de Medicina da Universidade de *Sheffield*, estudantes de licenciatura de economia da saúde e de medicina da mesma universidade e doentes (Brazier *et al.*, 1998); (v) os 611 indivíduos entrevistados no estudo de valoração do SF-6D no Reino Unido representam 0,001% da população total do Reino Unido, tendo os indivíduos deste estudo o mesmo peso na população total de Portugal (0,001%); (vi) ao contrário da investigação de Brazier, Roberts e Deverill (2002), foram tomados cuidados na selecção da amostra de indivíduos e no cálculo da dimensão da amostra, por forma a assegurar a utilização de um método de sondagem probabilístico que permitisse fazer inferências para a população portuguesa e conhecer as margens de erro dos resultados alcançados.

A inclusão do termo PIOR representando as interações não trouxe uma verdadeira melhoria aos modelos,

pelo que trabalhos futuros deverão centrar-se sobre a importância e o papel das interações nos modelos subjacentes ao sistema de classificação SF-6D.

A forma funcional usada no modelo foi aditiva linear à semelhança do modelo subjacente ao EQ-5D (Dolan, 1997), dado a natureza das variáveis independentes. Porém, alguns autores (Stratmann-Schoene *et al.*, 2006), recomendam os modelos multiplicativos na estimação de funções de preferência para sistemas de classificação de estados de saúde multi-atributo. Quando comparados com os modelos aditivos, têm a capacidade de capturar as interações entre atributos até um determinado grau.

No que diz respeito à capacidade de previsão do modelo, as representações gráficas parecem indicar que os erros de previsão sistemáticos tendem a verificar-se nos estados de saúde mais graves, tal como se havia verificado no sistema de pesos britânico (Brazier, Roberts e Deverill, 2002).

No entanto, apesar das limitações apresentadas, este é, de facto, um passo em frente no sentido de obter sistemas de valores adaptados à realidade portuguesa, para instrumentos de medição da QVRS, baseados em preferências, como o SF-6D. Por outro lado, é de referir que a grande mais-valia deste estudo prende-se com o facto de ser um dos primeiros a determinar um sistema de valores para o SF-6D para um país que não o Reino Unido. Com efeito, estão neste momento a ser conduzidos estudos de valoração para o SF-6D no Japão, Hong-Kong, Austrália e Brasil, mas apenas o do Japão se encontra numa fase mais avançada; os restantes ainda se encontram numa fase embrionária.

Por último, importa referir que alguns autores (Huang *et al.*, 2007) realçam ainda a necessidade da validação do sistema de valores obtido para um determinado país a partir da população em geral, em amostras de doentes e em subgrupos culturais eventualmente existentes na população em geral. A validação do sistema português de valores para o SF-6D apresentado neste artigo já foi iniciada em amostras de indivíduos com algumas doenças com elevada prevalência em Portugal, como a artrite reumatóide (Ferreira *et al.*, 2008) e as cataratas (Ferreira e Ferreira, 2009) e está neste momento a ser desenvolvida noutras doenças também com elevada prevalência em Portugal, como a asma e a doença pulmonar obstrutiva crónica. Prevê-se para um futuro próximo a sua validação em oncologia.

Agradecimentos

Lara N. Ferreira e Luís N. Pereira são beneficiários de bolsas de investigação para doutoramento (SFRH/

BD/25697/2005 e SFRH/BD/36764/2007, respectivamente), da Fundação para a Ciência e a Tecnologia. Versões prévias deste trabalho foram apresentadas na 14th Annual Conference of the International Society for Quality of Life Research, em Toronto, Canadá (Outubro de 2007), nas XV Jornadas de Classificação e Análise de Dados, em Setúbal, Portugal (Fevereiro de 2008) e na 7th European Conference on Health Economics, em Roma, Itália (Julho de 2008).

□ Referências bibliográficas

- BADIA, X. *et al.* — A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making*. 21 : 1 (2001) 7-16.
- BRAZIER, J. *et al.* — Measuring and valuing health benefits for economic evaluation. Oxford : Oxford University Press, 2007.
- BRAZIER, J.; ROBERTS, J.; DEVERILL, M. — The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*. 21 : 2 (2002) 271-292.
- BRAZIER, J. *et al.* — Estimating a preference-based index for a menopause specific health quality of life questionnaire. [Em linha]. *Health and Quality of Life Outcomes*. 3 : 13 (2005). [Consult. 2008-02-02]. Disponível em <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=15769294>
- BRAZIER, J. *et al.* — Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology*. 51 : 11 (1998) 1115-1128.
- BULLINGER, M. *et al.* — Translating health status questionnaires and evaluating their quality : the IQOLA project approach. *Journal of Clinical Epidemiology*. 51 : 11 (1998) 913-923.
- BUSSCHBACH, J. *et al.* — A comparison of EQ-5D time trade-off values obtained in Germany, the United Kingdom, Spain. In BROOKS, R., RABIN, R., CHARRO, F.D., ed. lit. — The measurement and valuation of health status using EQ-5D : a European perspective. Boston, MA : Kluwer Academic Publishers, 2003.
- CRAIU, R.; DUCHESNE, T.; FORTIN D. — Inference methods for the conditional logistic regression model with longitudinal data. *Biometrical Journal*. 50 : 1 (2008) 97-109.
- CUI, J.; QIAN, G. — Selection of working correlation structure and best model in GEE analyses of longitudinal data. *Communications in Statistics: Simulation and Computation*. 36 : 5 (2007) 987-996.
- DOLAN, P. — Modelling valuations for EuroQol health states. *Medical Care*. 35 : 11 (1997) 1095-1108.
- FERREIRA P. L.; SANTANA P. — Percepção de estado de saúde e de qualidade de vida da população activa : contributo para a definição de normas portuguesas. *Revista Portuguesa de Saúde Pública*. 21 : 2 (2003) 15-30.
- FERREIRA, L.; FERREIRA, P. — Qualidade de vida em doentes com cataratas. *Acta Oftalmológica (no prelo)*. (2009).
- FERREIRA, L.; FERREIRA, P.; BALEIRO, R. — Qualidade de vida em doentes com artrite reumatóide. *Acta Reumatológica Portuguesa*. 33 : 3 (2008) 341-342.
- FERREIRA, P. L. — Criação da versão portuguesa do MOS SF-36 : Parte I — Adaptação cultural e linguística. *Acta Médica Portuguesa*. 13 (2000a) 55-66.
- FERREIRA, P. L. — Criação da versão portuguesa do MOS SF-36 : Parte II — Testes de validação. *Acta Médica Portuguesa*. 13 (2000b) 119-127.
- FERREIRA, P.; FERREIRA, L. — A medição de preferências em saúde na população portuguesa. *Revista Portuguesa de Saúde Pública*. 24 : 2 (2006) 5-14.
- FURLONG, W. *et al.* — Guide to design and development health-state utility instrumentation. Ontario, Canada: McMaster University Press, 1990. (Working Paper Series).
- GANDEK, B.; WARE, J. — Methods for validating and norming translations of health status questionnaires : the IQOLA Project approach. *Journal of Clinical Epidemiology*. 51 : 11 (1998) 953-959.
- GREINER, W. *et al.* — A single European currency for EQ-5D health states : results from a six-country study. *European Journal of Health Economics*. 4 : 3 (2003) 222-231.
- HARDIN, J.; HILBE, J. — Generalized estimating equations. Boca Raton : Chapman & Hall/CRC, 2003.
- HAWTHORNE, G. *et al.* — The SF36 version 2 : critical analyses of population weights, scoring algorithms and population norms. *Quality of Life Research*. 16 : 4 (2007) 661-673.
- HUANG, I.-C. *et al.* — US and UK version of the EQ-5D preference weights : does choice of preference weights make a difference? *Quality of Life Research*. 16 : 6 (2007) 1065-1072.
- JOHNSON, J. *et al.* — Valuation of EuroQol (EQ-5D) health states in an adult US sample. *Pharmacoeconomics*. 13 : 4 (1998) 421-433.
- JOHNSON, J. *et al.* — Valuations of EQ-5D Health States : are the United States and United Kingdom different? *Medical Care*. 43 : 3 (2005) 221-228.
- JOHNSON, J. *et al.* — Comparison of Finnish and US-based visual analog scale valuations of the EQ-5D measure. *Medical Decision Making*. 20 : 3 (2000) 281-289.
- JONES, A. — Applied econometrics for health economists : a practical guide. Oxon, UK : Radcliffe Publishing, Office of Health Economics, 2007.
- JONES, A. *et al.* — Applied health economics. Great Britain : Routledge Taylor & Francis Group, 2007.
- LE GALÈS, C. *et al.* — Development of preference-weighted health status classification system in France : the health utilities index 3. *Health Care Management Science*. 5 : 1 (2002) 41-51.
- LIANG, K.-Y.; ZEGER, S. — Longitudinal data analysis using generalized linear models. *Biometrika*. 73 : 1 (1986) 13-22.
- MCCABE, C. *et al.* — Health state values for the HUI2 descriptive system : results from a UK survey. *Health Economics*. 14 : 3 (2005) 231-244.
- PAN, W. — Akaike's information criterion in generalized estimating equations. *Biometrics*. 57 : 1 (2001) 120-125.
- PATRICK, D. *et al.* — Measuring preferences for health states worse than death. *Medical Decision Making*. 14 : 1 (1994) 9-18.
- PORTUGAL. INSTITUTO NACIONAL DE ESTATÍSTICA. INE. — Recenseamento da população e da habitação (Portugal) : censos 2001. Lisboa : Instituto Nacional de Estatística, 2001.
- RABE-HESKETH, S.; EVERITT B. — A handbook of statistical analyses using stata. Boca Raton : Chapman & Hall/CRC, 2007.

SÄRNDAL, C. E.; SWENSSON, B.; WRETMAN, J. — Model assisted survey sampling. New-York : Springer-Verlag, 1992.

STRATMANN-SCHOENE, D. *et al.* — A preference-based index for the SF-12. *Health Economics*. 15 : 6 (2006) 553-564.

SZABO, S. *et al.* — An approach to response scale development for cross-cultural questionnaires. *European Psychologist*. 2 : 3 (1997) 270-276.

SZENDE, A.; OPPE, M.; CHARRO, F.D. — Comparative review of time trade-off value sets. In SZENDE, A., OPPE, M., DEVLIN, N. ed. lit. — EQ-5D value sets : inventory, comparative review and user guide. EuroQol Group Monographs, 2007. 21-28.

THE WHOQOL GROUP. — The World Health Organization Quality of Life Assessment (WHOQOL) : position paper from the World Health Organization. *Social Science & Medicine*. 41 : 10 (1995) 1403-1409.

TORRANCE, G. — Measurement of health state utilities for economic appraisal. *Journal of Health Economics*. 5 : 1 (1986) 1-30.

ZEGER, S. L.; LIANG, K.-Y. — Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 42 : 1 (1986) 121-130.

□ Abstract

THE PORTUGUESE SYSTEM OF WEIGHTS FOR THE SF-6D

Introduction: There has been an increase of surveys in some countries to obtain valuation functions for the most used preference-based instruments. Evidence suggests that health state valuations may differ from country to country. This study

seeks to define the Portuguese system of weights for the SF-6D.

Methods: A sample of 55 health states defined by the SF-6D has been valued by a random stratified sample of the general population using the standard gamble (SG). The survey was conducted by personal interview. In the SG valuation task, respondents were asked to value six health states, plus the worst health defined by the SF-6D or immediate death. Several models were estimated aiming to predict health state valuations. Firstly, linear models were estimated at both the individual and aggregate levels, as well as models with fixed effects, main effects, with interaction effects and with the constant forced to unity. Random effects models (REM) were also estimated using generalized least squares regressions. Generalized estimation equations (GEE) were used to estimate the REM with the constant forced to unity. Alternative functional forms were considered to account for the skewed distribution of the health state valuations. The models were analyzed in terms of their coefficients, overall fit and the ability for predicting the SG-values.

Results: The REM and the GEE models produced significant coefficients, which were found to be robust across model specification. The use of a battery of goodness of fit measures lead to the conclusion that the GEE model with main effects was the best model estimated.

Conclusion: This research demonstrates that it is possible to estimate preference weights for health related quality of life measurement. The best model seems to adequately predict the health states values of the Portuguese general population. Nevertheless, there are concerns regarding some inconsistent estimates that need further research.

Keywords: health state preference-based data; generalized estimation equations; random effects models; health related quality of life; SF-6D; health state utilities.

**ANÚNCIO
ASPECTOS GERAIS DE
TOXICOLOGIA**

p. 26 da revista 2-2008