ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/vhri

Methodology

A Bayesian Nonparametric Approach for Modeling SF-6D Health State Utility Scores

Samer A. Kharroubi, PhD, CStat

ABSTRACT

Objectives: Typically, models that were used for health state valuation data have been parametric. Recently, many researchers have explored the use of nonparametric Bayesian methods in this field. In this article, we report on the results from using a nonparametric model to predict a Bayesian short-form 6-dimension (SF-6D) health state valuation algorithm along with estimating the effect of the individual characteristics on health state valuations.

Methods: A sample of 126 Lebanese members from the American University of Beirut valued 49 SF-6D health states using the standard gamble technique. Results from applying the nonparametric model were reported and compared with those obtained using a standard parametric model. The covariates' effect on health state valuations was also reported.

Results: The nonparametric Bayesian model was found to perform better than the parametric model at (1) predicting health state values within the full estimation data and in an out-of-sample validation in terms of mean predictions, root mean squared error, and the patterns of standardized residuals and (2) allowing for the covariates' effect to vary by health state. The findings also suggest a potential age effect with some gender effect.

Conclusions: The nonparametric model is theoretically more flexible and produces better utility predictions from the SF-6D than previously used classical parametric model. In addition, the Bayesian model is more appropriate to account the covariates' effect. Further research is encouraged.

Keywords: covariates, nonparametric Bayesian methods, preference-based health measure, SF-6D, standard gamble.

VALUE HEALTH REG ISSUES. 2022; 27:1-11

Introduction

A large number of preference-based measures of health-related quality of life available have been used to measure quality-adjusted life years (QALYs) to be applied in cost-effectiveness analyses. Such measures include the EQ-5D,[1] Health Utilities Index (HUI) 2 and 3,[2,3] Assessment of Quality of Life,[4] Quality of Well Being scale,[5] and the short-form 6-dimension (SF-6D) health index,[6] in addition to a growing number of condition-specific preference-based measures.[7,8] A key practical issue with these measures is that they produce lots of distinct health states, so it is not practical to get direct valuation for every health state. Hence, models are then needed to estimate utilities for all states defined by the measure based on the valuations of a representative sample of states.

Typically, the distribution of health state values obtained from valuation techniques such as time trade-off and standard gamble (SG) is skewed, truncated, hierarchal, and non-continuous, all of which impose a key challenge for statistical modeling methods.[6] Despite this challenge, earlier statistical methods of such data have achieved some success in the SF-6D,[6] EQ-5D,[9] and HUI2[10]; however, there were some concerns regarding the size of the prediction errors and an obvious systematic pattern associated with these prediction errors. Besides, these models are limited in how they capture the covariates' effect on health state values.

To overcome these challenges, an alternative nonparametric Bayesian method was reported for modeling health state preference valuations.[11] This method offers more flexible and realistic inference than previously used standard parametric models,[6,9,10,12,13] in addition to being more flexible in capturing the covariates' effect. Kharroubi et al[14] have used this method of modeling for SF-6D UK valuation data and expanded it to deal with covariates.[15] Furthermore, Kharroubi and McCabe[16] and Kharroubi and Abou Daher[17] have used it to model other health state measures such as the UK HUI-2 and EQ-5D and for other countries such as SF-6D Hong Kong and Japan.[18,19] Furthermore, it has been extended to handle the joint US/UK EQ-5D valuation data[20] and the joint UK/Hong Kong and UK/Japan SF-6D data.[21,22] This article extends the method to the SF-6D Lebanon valuation data and compares the findings with the previously applied standard random effect model. Despite this article not presenting new methodological advances because the models used are replications of the methods used to model the UK data, it tells a reassuring story about the superiority and the better performance

of the nonparametric Bayesian approach in generating accurate estimates compared with the parametric approach.

The Methods section of this article briefly presents the SF-6D Lebanon valuation study along with the data collection procedure. Full details of the valuation study and data collection process are available elsewhere.[23] The parametric and nonparametric methods for health state valuations are also outlined. The findings obtained from every method while comparing the models in terms of their predictive ability are set out in the Results section. The ability of the nonparametric Bayesian model to capture the effect of the covariates is also set out. The Discussion section concludes with a general discussion of the findings along with some limitations for the study and possible directions for further research.

Methods

The SF-6D

The SF-6D is a generic measure of health states derived from the 36-Item Short Form Survey. This preference-based measure includes 6 dimensions, including physical functioning, role limitations, social functioning, pain, mental health, and vitality, each having between 4 and 6 levels.[6] Health states of the SF-6D are represented by a 6-digit number by choosing a level from every dimension, beginning with physical functioning and finishing off with vitality, generating 18 000 possible combinations in total. The best health state is expressed as 111111, which is the perfect health state indicating no loss of health in any dimension, whereas the descriptor 645655 represents the worst health state known as "the pits."

Study Design

A sample of 126 Lebanese respondents from the American University of Beirut valued 49 health states described through the SF-6D using the SG technique. Selection of respondents along with selection of health states has been reported elsewhere.[23]

Selection of respondents

Lebanese adults aged 18 years or older were recruited at the American University of Beirut and were stratified by sex (male or female) and profession (faculty, staff, or students). Respondents were contacted by phone or email to participate in the study. Respondents were excluded if they could not be reached after 2 attempts or if they refused to participate.

A formal sample size calculation was not undertaken. Nevertheless, previous experience with the SF-6D in many countries, such as the UK, Hong Kong, and Japan, and more recent work in valuing the EQ-5D has demonstrated that 15 observations per health state are sufficient to estimate an additive model. Hence, a total of 126 people with 21 in each of the 6 gender-occupation groups were recruited from 170 initially contacted respondents giving a response rate of 74%. The reasons for declining to participate were mainly time constraints and disinterest in the study. Each 1 of the 7 sets of health states was valued by 3 respondents from every group (gender-occupation) with a total of 18 valuations per health state. More on the selection method is given in the study by Kharroubi et al.[23]

Selection of health states

The sample of health states was chosen using the Orthoplan procedure in SPSS (SPSS Inc, Chicago, IL), which generated 49 health states of the 18 000 possible states required to fit an additive model. The 49 states were chosen at random in the sense that every level of each dimension was included at least once, different combinations of dimension levels in different states, and the combination of states already existed in available SF-6D data sets. These 49 health states were distributed in over 7 sets with 7 health states in every set. Additionally, all respondents valued the "pits" state. More details are available in the study by Kharroubi et al.[23]

Interviews

This study used the same interview protocol and schedule as the UK study. Each respondent was asked to rank and then value 8 SF-6D health states using the McMaster "ping pong" variant of SG, where the respondent is asked to value each of the 7 certain health states against the full and "pits" health state.[24] As for the eighth SG question, each respondent was asked to value the "pits" state. On the basis of their valuation of the "pits," each respondent was asked to choose between either (1) the certain prospect of being in the "pits" state and the uncertain prospect of full health or immediate death or (2) the certain prospect of death and the uncertain prospect of full health or the "pits" state. The chances that the best outcome occurring are varied until the respondent is indifferent between the 2 prospects. Any negative value was bounded to −1 and it referred to worse than death.[25] The 7 earlier health state valuations (SG) were adjusted onto the scale of 1 to 0, with 1 being the perfect health state and 0 indicating death,[6] using the following formula:

$$\mathrm{SGADJ} = \mathrm{SG} + (1 - \mathrm{SG}) \times P,$$

where P is the valuation of the "pits" state. The SGADJ values represent the response variable (y) in the 2 models described below.

Study Sample

As already mentioned, each respondent has been asked to value 8 SF-6D health states including "pits." This results in 1008 health state valuations across 49 health states. There were 2 respondents whose valuations did not change among the 8 health states and so they had to be excluded from the analysis, leaving in total 992 (98.4%) SG valuations across 49 health states for the analysis. Every health state has been valued 17 or 18 times, with average values ranging from 0.322 (pits state) to 0.890 (state 211111) and standard deviations ranging from 0.042 (211111) to 0.265 (614434). Nevertheless, the median values were higher, indicating the left skewness of the data. Negative values did not occur and more than 19% of observations were above 0.9. Finally, no states were evaluated at 1.0, indicating that all respondents were willing to risk a worse state to get a chance for a better state. More details on health state values are available in Kharroubi et al.[23]

Modeling

Valuation surveys typically produce data with a complex structure, like skewness or heteroscedasticity which creates some problems for conventional statistical modeling procedures. A variety of techniques have been adopted by Dolan,[9] Brazier et al,[6] McCabe et al,[10] and Shaw et al[13] to address these problems using parametric methods and by Kharroubi et al[11,14-17] using a nonparametric method. We briefly discuss the 2 methods here.

The Parametric Approach

A general parametric model for individual j's valuation of health state i can be specified as follows:

$$yij = \mu + \theta' g(xij, tj) + \alpha j + \varepsilon ij \quad (1)$$

where, for i = 1,2,…n_j and j = 1,2,…m; x_{ij} is the ith health state valued by individual j; and y_{ij} is the SGADJ value obtained by individual j for that health state. Furthermore, μ and θ represent unknown parameters, $g(x_{ij},t_j)$ represents a vector of regressors, and t_j denotes individual characteristics such as age and sex or socioeconomic factors of respondents. The model also includes a zero-mean random terms: the individual effect term α_j of respondent j and the error term ϵ_{ij}.

Every dimension of x_{ij} was treated as a factor; thus, $g(x_{ij},t_j)$ includes a dummy indicator variable for every level λ above 1 of each dimension δ of the SF-6D. For instance, x_{52} represents dimension δ = 5 (mental health) and level λ = 2 (you feel tense or downhearted and low a little of the time). Thus, for such a health state, $x_{\delta\lambda}$ takes the value 1 when dimension δ is at level λ and 0 otherwise.

In all, a total of 25 regressor variables is defined, where level λ = 1 of every dimension is treated as a baseline. Therefore, the intercept parameter μ denotes the perfect state 111111, and adding the parameters θ of the "on" dummies to μ attains the value of any other health state. Note that $g(x_{ij},t_j)$ could also contain other dummy variables to take into account the interactions among the levels of different dimensions, but we do not consider interaction adjustment here.

Finally, generalized least squares method was used to estimate this random effects model. Given the mean of α_j is zero, the population health state utility for a given state x is derived as $\mu + \theta' g(x,t)$.

The Nonparametric Approach

Models, such as (1), have been parametric, and, therefore, they allow the utility function to take a particular form. Alternatively, Kharroubi et al[11] proposed a nonparametric approach that allows this utility function to take any form at all, employing Bayesian hierarchical modeling. Kharroubi et al[14] used this model on the UK SF-6D valuation data, and Kharroubi et al[15] expanded the model to address covariates. Kharroubi and McCabe[16] and Kharroubi and Abou Daher[17] estimated HUI 2 and EQ-5D valuations data, respectively, using the same model. Here, we follow on to apply this model to the Lebanon SF-6D data set.

The Bayesian statistical nonparametric model describes the characteristics of individual health state valuation data in such a way that the ith valuation by individual j is modeled as follows:

$$yij = 1 - \alpha j\{1 - u(\mathrm{xij})\} + \varepsilon ij, \tag{2}$$

where the individual residual and error terms are assigned the following distributions:

$$\alpha j \sim LN(tjT\gamma, \tau 2) \text{ and } \varepsilon ij \sim N(0, \nu 2), \tag{3}$$

t_j denotes a vector of covariates for individual j and γ, τ^2, and ν^2 are unknown parameters. It follows from (2) that the random effect and covariates enter multiplicatively, whereas in (1) they enter additively, thereby allowing for more effect the further the base utility is away from 1. Therefore, the variation in individuals' valuations will be greater for health states with lower utility than for those with higher utility. For more details on this, see Kharroubi et al.[11]

From the definition of the model (2), the term $u(x)$ is an unknown function and represents the median utility for health state x. Therefore, in Bayesian inference, it will be treated as a random variable. The previous distribution of u(x) is defined as

$$u(x)|\beta, \sigma 2 \sim N(\mathrm{h(x)T}\beta, \sigma 2) \tag{4}$$

where $h(x) = (1, x)^T$, x being the vector of discrete levels on each of the 6 health state dimensions, and β and σ^2 denotes are unknown parameters. It is to be noted that the conditional expectation of (4) expresses a belief that $u(x)$ will be roughly linear and additive in the various dimensions, whereby in (1) the assumptions of linearity and additivity are imposed. The actual utility is free to vary around this mean, as per its multivariate normal distribution, and therefore, it may take any form at all. It is in this sense model (2) is described as nonparametric, and it is believed that this is another way in which model (2) is more realistic than that of (1). See Kharroubi et al[11] for thorough details about this part of the model.

Furthermore, there seems to be a high correlation between $u(x)$ and $u(\mathbf{x}')$ when the health states x and $\mathbf{x}'$ get sufficiently close, and according to Kharroubi,[11] this is defined as

$$\mathrm{cov}(u(x), u(x')|\sigma 2) = \sigma 2\ c(x, x'), \tag{5}$$

where

$$c(x, x') = \exp\left\{-\sum bd(xd - x_d)2\right\} \tag{6}$$

where for d = 1, 2, …, 5, x_d and x'_d represent the levels of dimension d in x and $\mathbf{x}'$, respectively, and b_d is the roughness parameter that controls the degree of adherence of $u(x)$ to a linear form in dimension d. For more explanation about this particular point, see Kharroubi et al.[11]

To this end, it follows from Kharroubi et al[11] that the mean health state utility in (2) is

$$\tilde{u}(x) = 1 - E(\alpha)\{1 - u(x)\} \tag{7}$$

where $E(\alpha)$ denotes the expected value of α over the whole population. Note that, if $E(\alpha)$ = 1, then $\tilde{u}(x) = u(x)$, which is not generally the case. Kharroubi et al[11] provide more explanation on the computation of $E(\alpha)$.

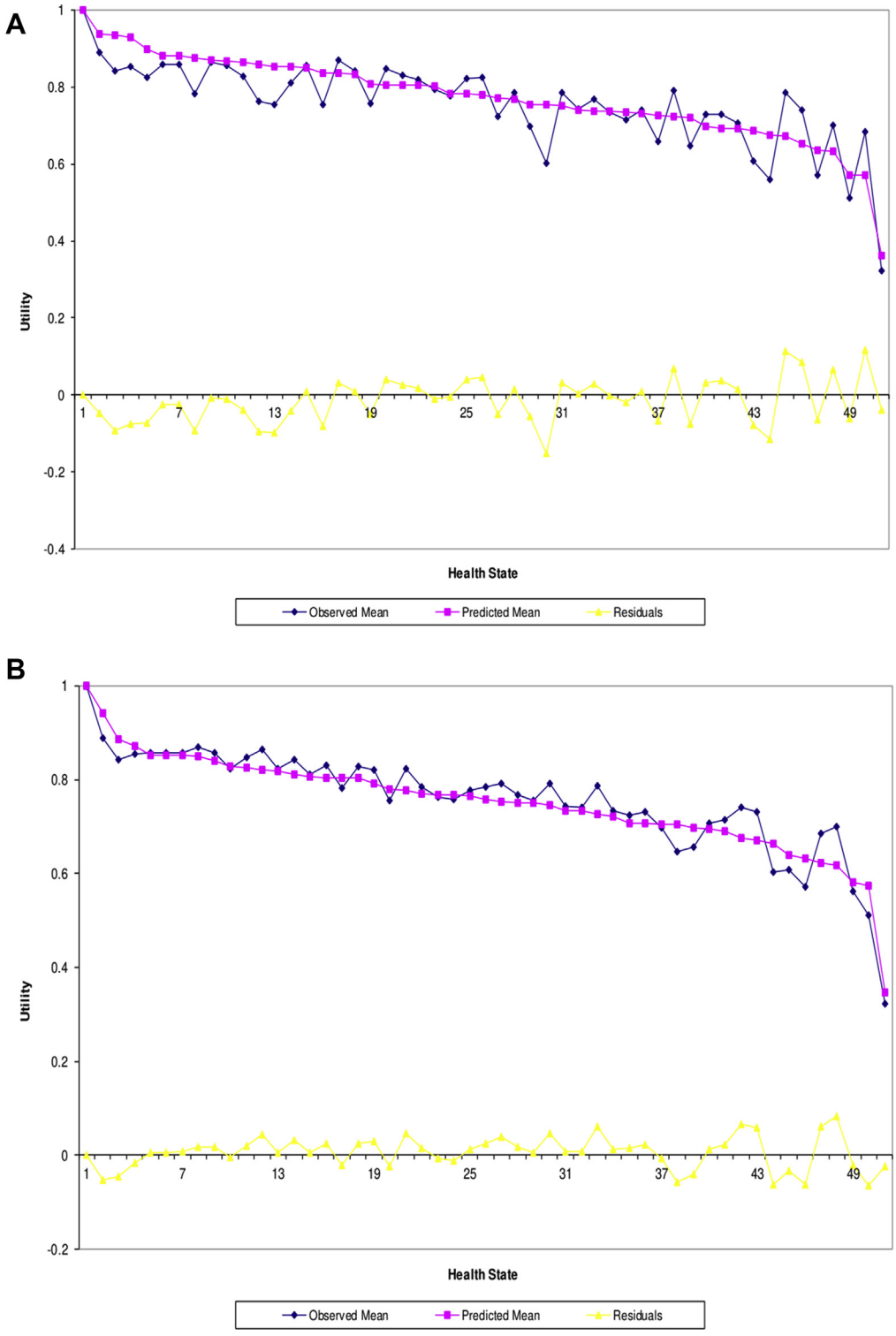
Programs to undertake models (1) and (2) were written in MATLAB, and the relevant code can be found at https://www.sheffield.ac.uk/scharr/sections/heds/mvh/sf-6d/bayesian. Alternatively, an Excel spreadsheet that produces estimated utilities and standard deviations for health states described by the SF-6D instruments is also available on the same website for ease of use in practice.

Results

No Covariates

For this analysis, covariate were not included so we set $\ln(\alpha_j) \sim N(0, \tau^2)$, and following Kharroubi et al,[11] we let $h(x) = (1, x_1, x_2, x_3, x_4, x_5, x_6)$ and $b_d = 2.5/(l_d - 1)^2$, where l_d represents the number of levels in dimension d. The predictive performance of the 2 models is compared in Figure1A,B, where the predicted (line marked with squares) and observed (line marked with diamonds) mean valuations for the 49 health states valued in the sample were displayed for models (1) and (2), respectively. In each figure, the errors are represented by the line marked with triangles and are calculated by taking the difference between the 2 valuations. It is to be noted that health states were sorted by predicted mean valuations and plotted accordingly. When comparing the plots, Figure 1A shows an obvious variation of the observed values around the predicted mean valuations and also reveals a tendency to overpredict at high (ie, good health states) and intermediate (ie, moderate health states) health state values. In contrast, Figure 1B suggests that model (2) predicts the mean utilities quite well across the board,

Figure 1. Actual and predicted mean health states valuations for the (A) parametric model and (B) nonparametric model.



with a slight underprediction for the good states. This provides an initial indication that model (2) is less prone to systematic bias.

Table 1 presents the inference for the mean health state valuations of the 49 health states used in the valuation survey. For each state, Table 1 reports the posterior mean and standard deviation from both models (1) and (2) along with the actual mean utility. Across the 49 states, the predictive performance of model (2) is better than that of model (1) overall, with a root mean square error (RMSE) of 0.035 for model (2) compared with 0.061 for model (1). Additionally, no health state is predicted as being worse than death for both models. Nevertheless, Table 1 shows potential differences between the 2 models. For example, the

Table 1. Inference for the 49 health states.

Health state	N	Observed mean	Nonparametric posterior inference (no covariates)		Parametric posterior inference		Nonparametric posterior inference (covariates)	
			Mean	SD	Mean	SD	Mean	SD
111621	18	0.824	0.828	0.026	0.898	0.044	0.828	0.026
113411	18	0.854	0.871	0.021	0.929	0.046	0.871	0.021
115653	18	0.731	0.707	0.027	0.699	0.048	0.706	0.027
121212	18	0.842	0.887	0.019	0.934	0.047	0.887	0.020
122233	17	0.869	0.851	0.024	0.837	0.052	0.851	0.023
122425	17	0.758	0.769	0.023	0.808	0.050	0.768	0.024
124125	18	0.848	0.827	0.025	0.807	0.048	0.828	0.024
131542	18	0.828	0.804	0.024	0.866	0.047	0.804	0.024
132524	18	0.763	0.769	0.022	0.859	0.050	0.769	0.022
133132	18	0.858	0.852	0.023	0.883	0.048	0.851	0.023
135312	18	0.756	0.780	0.022	0.855	0.049	0.779	0.022
142154	17	0.791	0.753	0.029	0.724	0.049	0.753	0.029
144341	17	0.742	0.734	0.025	0.733	0.051	0.733	0.025
211111	18	0.890	0.943	0.017	0.939	0.043	0.944	0.017
212145	18	0.785	0.769	0.027	0.752	0.049	0.770	0.027
213323	18	0.783	0.805	0.021	0.875	0.051	0.804	0.021
221452	18	0.824	0.818	0.023	0.783	0.051	0.817	0.023
224612	18	0.646	0.704	0.024	0.722	0.053	0.704	0.024
232111	17	0.858	0.853	0.025	0.850	0.049	0.852	0.025
235224	17	0.768	0.751	0.024	0.737	0.055	0.750	0.024
241531	18	0.785	0.759	0.025	0.769	0.050	0.761	0.025
312332	18	0.864	0.821	0.023	0.870	0.049	0.820	0.023
315515	18	0.698	0.706	0.025	0.755	0.049	0.705	0.025
321122	18	0.858	0.853	0.022	0.882	0.048	0.853	0.022
323644	18	0.572	0.634	0.026	0.637	0.054	0.633	0.025
332411	17	0.844	0.813	0.025	0.834	0.051	0.812	0.025
334251	17	0.735	0.723	0.025	0.737	0.054	0.722	0.025
341123	18	0.831	0.806	0.025	0.806	0.050	0.807	0.025
412152	18	0.793	0.747	0.025	0.803	0.047	0.746	0.025
414522	18	0.756	0.751	0.023	0.837	0.052	0.749	0.023
421314	18	0.812	0.807	0.024	0.854	0.050	0.806	0.023
425131	18	0.658	0.697	0.024	0.725	0.051	0.697	0.023
431443	17	0.825	0.778	0.027	0.780	0.053	0.777	0.027
432621	17	0.743	0.734	0.024	0.740	0.051	0.733	0.024
443215	18	0.731	0.672	0.027	0.694	0.051	0.675	0.028
511114	18	0.858	0.841	0.025	0.869	0.045	0.840	0.025
512242	18	0.603	0.665	0.024	0.754	0.050	0.665	0.024
522321	18	0.778	0.766	0.023	0.783	0.050	0.765	0.023
523551	18	0.607	0.640	0.026	0.686	0.052	0.639	0.025
531635	17	0.787	0.726	0.030	0.673	0.051	0.726	0.030
534113	17	0.724	0.708	0.025	0.773	0.052	0.707	0.025
545422	18	0.701	0.618	0.029	0.635	0.052	0.622	0.029
611221	18	0.821	0.792	0.025	0.805	0.047	0.792	0.026
614434	18	0.561	0.583	0.027	0.676	0.052	0.581	0.027

continued on next page

Table 1. Continued

Health state	N	Observed mean	Nonparametric posterior inference (no covariates)		Parametric posterior inference		Nonparametric posterior inference (covariates)	
			Mean	SD	Mean	SD	Mean	SD
622513	18	0.707	0.695	0.026	0.692	0.050	0.694	0.026
625141	18	0.511	0.576	0.027	0.572	0.051	0.575	0.027
631355	17	0.741	0.675	0.031	0.654	0.051	0.674	0.031
633122	17	0.714	0.691	0.025	0.734	0.050	0.690	0.025
642612	18	0.685	0.623	0.029	0.570	0.048	0.626	0.028
645655	124	0.322	0.346	0.025	0.361	0.027	0.347	0.025

SD indicates standard deviation.

predicted mean utility value for the pits state is 0.361 for the parametric model compared with 0.346 for the nonparametric model, whereas the actual mean for this state is 0.322. Furthermore, Table 1 shows that the standard deviations of the estimates are bigger for model (1) given these are based on the assumption that the preference function is additive in the different factors. In contrast, the posterior standard deviations are smaller mainly because they come from a model that permits uncertainty about the shape of the utility function.

Now we assess the benefits in terms of bias and/or precision; Figure 2A,B shows the Bland–Altman agreement plots[26] for models (1) and (2), respectively. Here, the difference between the predicted and actual mean health state valuations is plotted versus the average of these valuations. In both figures, the solid line indicates the mean of the differences and the dotted lines are the 95% limits of agreement. When comparing the plots, it is apparent that 95% of the errors lie between the limits of agreement, with 3 extreme observations for model (2) and only 1 outlier for model (1). The plots also suggest that model (2) provides better agreement because the 95% limits width is narrower, with a value of 0.136 compared with 0.231 for model (1). Although the 2 models produce mean bias close to zero (0.006 vs −0.016), the standard deviation of the difference scores for the nonparametric model is less than the parametric, with values of 0.035 and 0.059, respectively.

Another aspect to show the difference between the parametric and nonparametric models is reflected in Figure 3, which shows the predicted values from the parametric model (Fig. 3A) and from the nonparametric model (Fig. 3B) versus the actual values of the 49 health states, along with the 45° line of unity (solid line). In theory, the fitted values are expected to lie roughly on the unity line; that could be evaluated as a good agreement. In practice, the plots suggest that the valuations from the nonparametric model tend to be closer to the perfect line, whereas the valuations from the parametric model have a larger scatter, so estimates deviate largely from the theoretical line. Therefore, the nonparametric model generates better predictions and much more precise estimates than the parametric one.

The validity of the 2 models is further verified in Figure 4 that presents histograms of standardized residuals across all 882 health state valuations for both the parametric (Fig. 4A) and nonparametric models (Fig. 4B). According to these models, these standardized residuals are expected to be approximately standard normally distributed, that is, $N(0, 1)$. This assumption is reinforced in Figure 4A,B, although there is a slight evidence of skewness that is more obvious in Figure 4A. The skewness is not surprising, given the negative skewness at the level of individual observations. Nevertheless, the analyses of the 2 models are not seriously invalidated, given that the skewness of the standardized residuals is probably not high enough to jeopardize them.

We now perform an out-of-sample leave-one-out prediction, where 10 health states were selected randomly from the

Figure 2. Bland–Altman plot for the predictive errors of the 42 health states for the (A) parametric model and (B) nonparametric model.

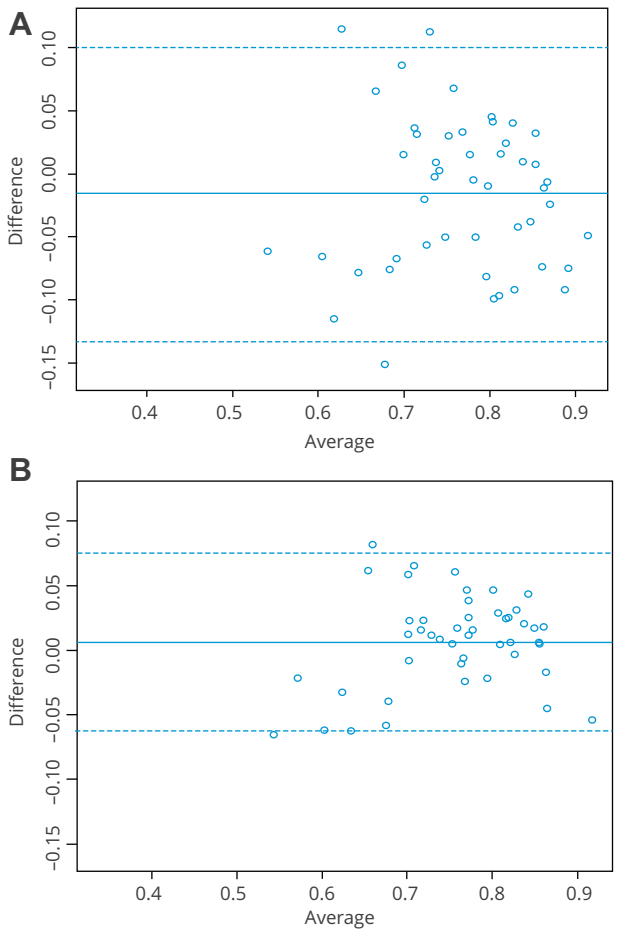
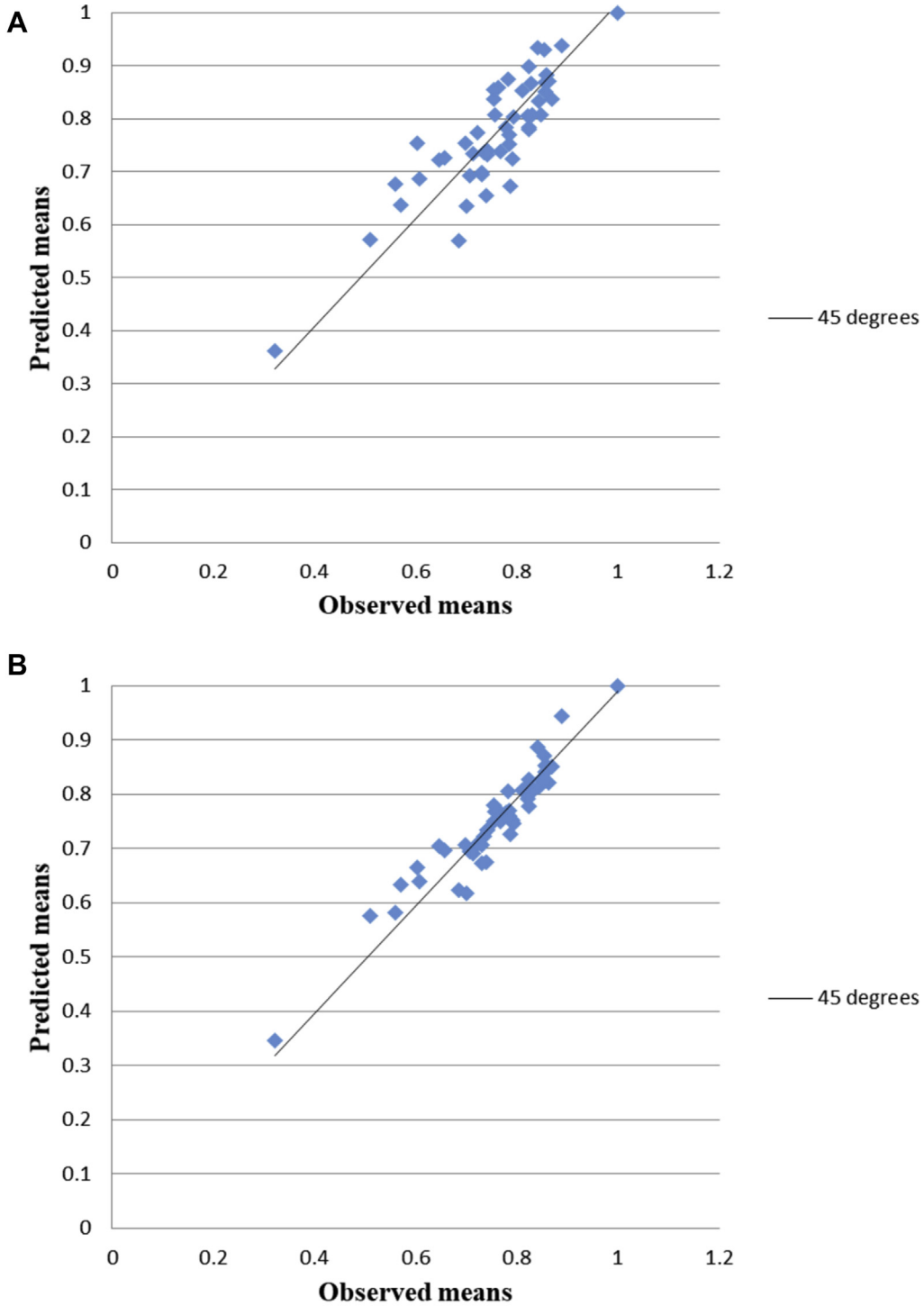


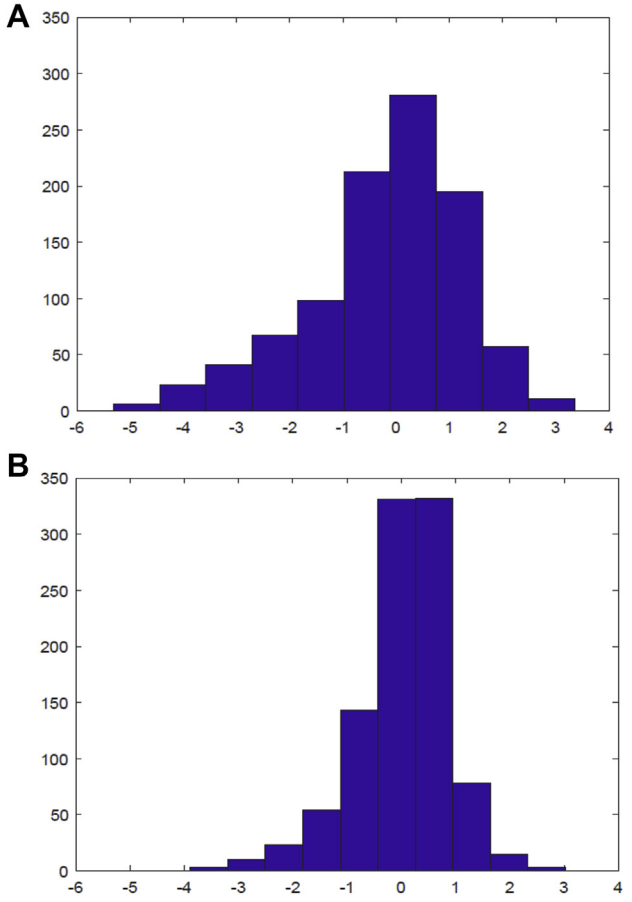
Figure 3. Actual and predicted mean health states valuations for the (A) parametric model and (B) nonparametric model.



estimated data; then consecutively valuations relating to every selected health state were omitted, and the models were fitted on the reduced data set of 39 health states. Table 2 presents the observed means for the 10 omitted states, along with their predicted means and standard error values from both models estimated on the reduced data set. Results revealed that the model (2) predicts the left-out data quite well and better than model (1) overall, with RMSE values of 0.054 and 0.099, respectively.

Figure 5A,B presents the Q-Q plots of standardized predictive errors for the 10 out-of-sample health states for models (1) and (2), respectively. The solid line in every case corresponds to the theoretical $N(0, 1)$ distribution and not fitted. Theoretically, the quantiles of the standardized predictive errors are expected to lie roughly on the solid line; hence, their distributions are identical. In Figure 5A, the dots tend to deviate substantially from the theoretical line, suggesting that model (1) may not be well

Figure 4. Standardized residuals of 882 individual health state valuations for the (A) parametric model and (B) nonparametric model.



validated by its predictive ability. In contrast, Figure 5B indicates that model (2) predictions are well validated. Almost identical results were also obtained across 5 replicates.

Covariate Effects

For this analysis, we include the covariates effects in the nonparametric model by setting $\alpha_j \sim LN\,(t_j^T\,\gamma,\,\tau^2)$ and $\mathbf{t_j}$ to be (age, age^2, sex). Note that t_j are centered, so they have zero means; as a result, the value of α_j for a typical person will be 1. To demonstrate the effect of accounting for sex and age on mean utilities, predicted mean and standard deviation values are reported for the health states reported in Table 2, with and without controlling for covariates. The actual Lebanese sex and age distributions were adapted from the Lebanese Central Administration of Statistics Lebanon of 2009.[27] It is apparent from Table 2 that the inclusion of the covariates resulted in almost identical posterior mean health state values to those that estimated the age and sex distributions from the valuation samples, with difference values of 0.005 or less in mean valuations.

The impact of these covariates is reflected in Figure 6A which presents the histograms of the conditional posterior distribution of sex, age, and age^2. The plots show that the posterior distributions for both age and age^2 are centered away from zero, thereby they have strong effects, whereas for sex the effect seemed to be weaker because the histogram is almost centered at zero. On that basis, the impact of age on health state values has been further investigated in Figure 6B, where the posterior mean utilities were plotted against age for the following 3 health states: the pits state (645655), a mild state (211111), and a moderate state (334251). For the worst state, it can be seen that the curve forms a clear quadratic shape, with the health state values that keep increasing until 55 to 60 years old, then gradually decreasing for the following 10 to 15 years, and then more sharply thereafter. Nevertheless, the mild and moderate states form a similar quadratic shape, but their shape appears to be less dramatic, especially for the mild one. Therefore, the impact of age depends heavily on how severe the health state in question is.

Discussion

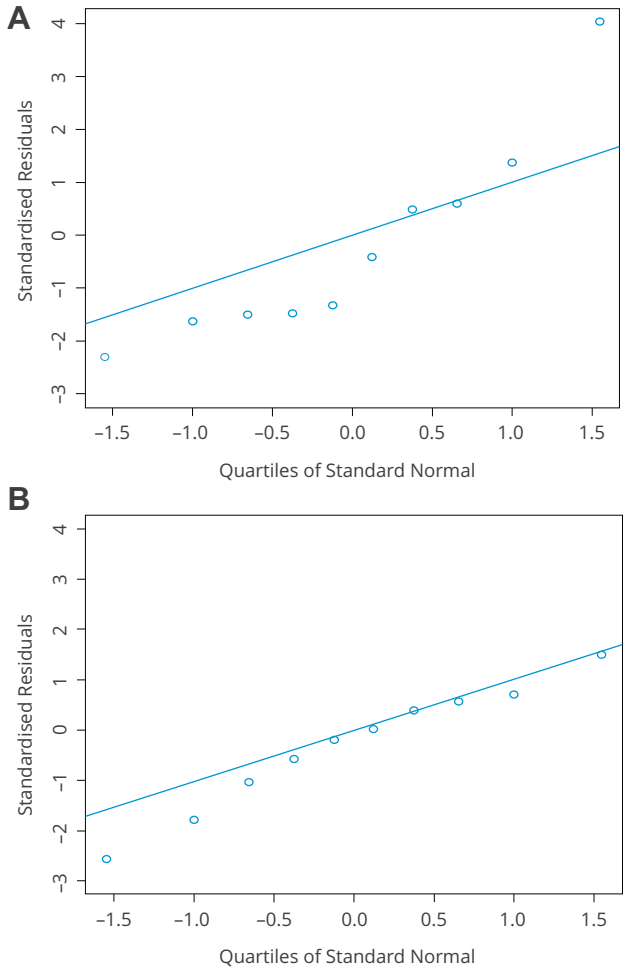
Here we have reported the findings from applying a nonparametric approach to modeling the SF-6D Lebanon valuation data using Bayesian methods. Two sets of analyses were

Table 2. Out-of-sample predictions for 10 health states.

Missing state	N	True sample mean	Nonparametric posterior inference		Parametric inference	
			Mean	SD	Mean	SD
121212	18	0.842	0.936	0.054	0.914	0.048
132524	18	0.763	0.738	0.056	0.834	0.053
211111	18	0.890	0.977	0.034	0.960	0.043
232111	17	0.858	0.912	0.050	0.833	0.051
321122	18	0.858	0.888	0.049	0.878	0.050
412152	18	0.793	0.715	0.050	0.763	0.051
432621	16	0.736	0.707	0.057	0.656	0.058
523551	18	0.607	0.560	0.071	0.701	0.062
614434	18	0.561	0.573	0.066	0.722	0.070
642612	18	0.685	0.679	0.062	0.482	0.050
RMSE			0.054		0.099	

RMSE indicates root mean square error; SD, standard deviation.

Figure 5. Q-Q plot of standardized predictive errors for the 10 of sample health states for the (A) parametric model and (B) nonparametric model.



presented; the first was comparing the nonparametric Bayesian model with the previously used standard parametric one. The second was using the flexibility of this model in capturing the impact of respondent characteristics on health state valuations.

It is clear that, from the analysis presented here, the predictive ability of the nonparametric Bayesian model is better than the previously used classical parametric model overall, as reflected in better RMSE within the full estimation data and better standardized predictive errors in the out-of-sample validation data. It may be argued that the improvement in RMSE is modest in this example; however, there are other potential advantages over and above the mean differences. More specifically, the Bayesian model is able to produce predictions that are much better behaved, a utility function with the key feature of capturing the lower (upper) utilities in the sample for the poorest (best) health states, and posterior variances and standardized predictive errors that correctly validate against the omitted sample. Furthermore, there are significant efficiencies that can be produced from survey design, data sampling, and ultimately data analysis if a nonparametric Bayesian method is employed by researchers.

The analysis of covariates in this study showed that age has a potential effect on health state values with some gender effect. The results also showed the impact of those variables depends heavily on the severity of the health state. More specifically, for milder health states (eg, 211111), the variability in the predicted values was quite minimal with respect to age. Nevertheless, as the health state becomes more severe, the impact of age on the health state value gets greater. The overall impact of covariates after controlling for covariates' age and sex was quite modest (typically 0.005 or less). Thus, even controlling for the impact of age (and sex), the distribution of the Lebanese general population had little impact on health state valuations. Analyses by Kharroubi at al,[14] Kharroubi and McCabe,[16] and Kharroubi and Abou Daher[17] of the UK SF-6D, EQ-5D, and HUI-2 valuation data have drawn similar conclusions. It would be interesting though to explore the performance of the nonparametric approach in settings where more covariate effects need to be handled. This is an important issue and is the subject of further work.

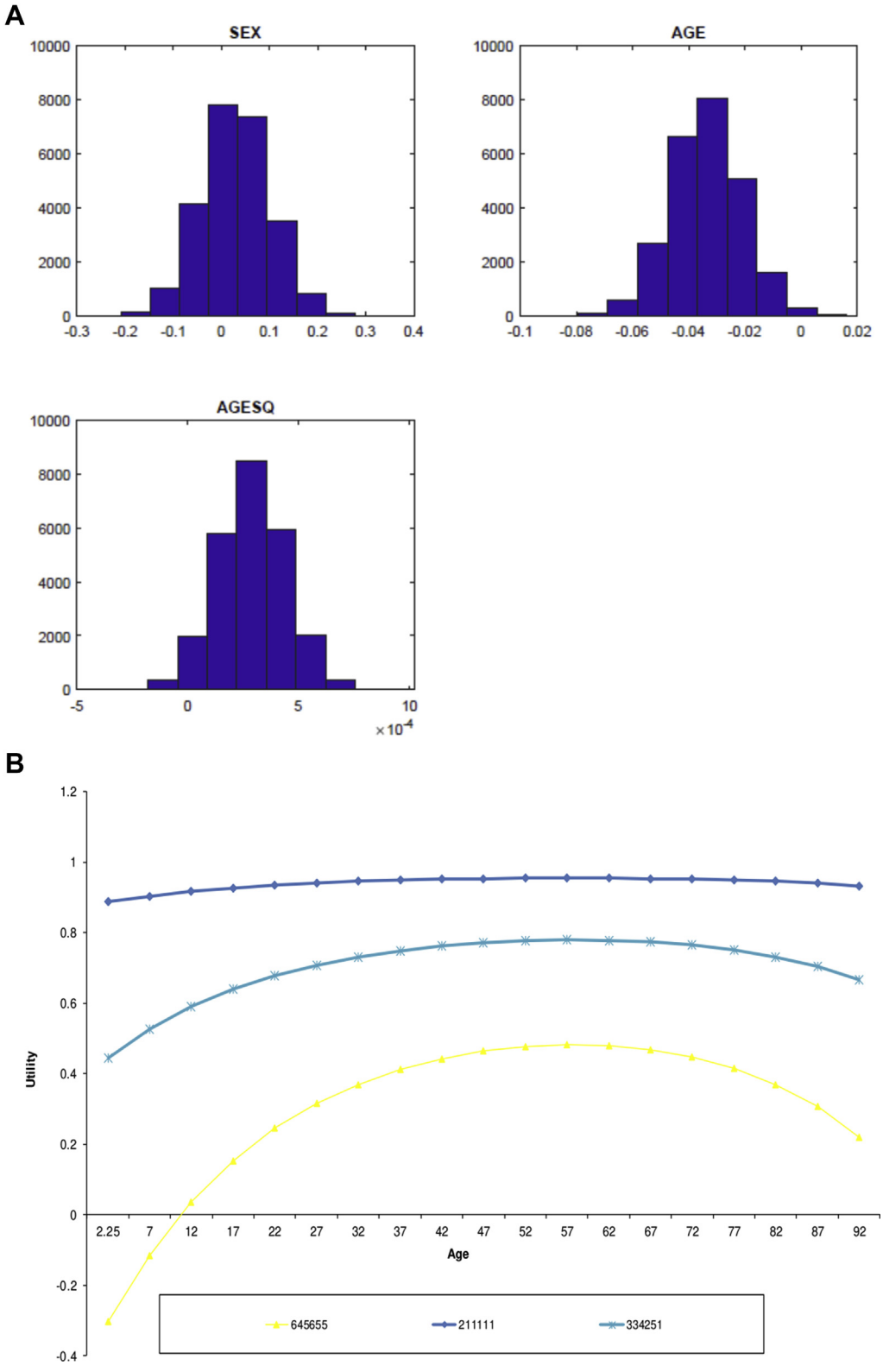
The use of the SF-6D has become widespread across the globe reaching the United States, the United Kingdom, Europe, and Asia and is largely available for use in data sets because it is derived from the commonly used 36-Item Short Form Survey. There is potential to borrow strength from existing countries' valuations to generate better representative utility estimates. The nonparametric model provides a key potential advantage because it allows making use of results of 1 country as informative priors to improve those of another country, and as such generated utility estimates of the second country will be much more precise than collecting and analyzing its data separately. This is a promising approach for other countries with smaller population size or low- and middle-income countries where the cost of undertaking large valuation surveys and collecting data via face-to-face interviews with techniques such as SG and/or time trade-off to generate country-specific value sets can be prohibitive, especially in the context of data collection after the coronavirus disease pandemic. Work is in progress to explore whether using the already existing UK value sets might contribute substantial informative prior for the analysis of the new valuation SF-6D study in Lebanon and preliminary results sound very promising.

To handle some of the problems set by the parametric approach, Méndez et al[28] have applied a semiparametric approach to modeling health state valuation data using inverse probability weighting techniques. This approach makes no assumption on the distribution of health state valuations, accommodates respondent characteristics, and allows for an undetermined amount of heterogeneity in the predictions. Further research is encouraged to compare the present nonparametric Bayesian approach with that used by Méndez et al[28] to make best use of health state valuations.

There are certain limitations to our study. First, the study sample was small, so this may somehow limit the generalizability of the preference values found. Second, the study sample was recruited from the American University of Beirut, and hence, the majority of our respondents were highly educated (88.7% hold a degree or above). The exclusion of respondents with low education levels could bias the valuation results because people from different socioeconomic backgrounds may value health differently, so they may give different utilities for the same health state. Third, the small number of health states valued could impact the accuracy of econometric modeling. Therefore, the SF-6D preference-based utilities from this pilot study should not be regarded as representative of the general Lebanese population. Future research with a larger and more representative sample from the general population of Lebanon is encouraged to generate the Lebanese specific SF-6D scoring algorithm.

An additional concern of the approach used is around negative utilities (ie, states worse than death). Although previous studies in many countries, such as the United Kingdom, Hong Kong, and Japan,[6,12,29] have observed a relatively low number of negative utilities elicited using the SG technique, the lack of negative

Figure 6. (A) Posterior distributions of main covariates. (B) Mean standard gamble utility values by age for 3 health states for the nonparametric model. AGESQ indicates age[2].



utilities is surprising in the current study. This could be caused by attitudes to risk, characteristics of the particular survey sample, and the perceived severity of the states. Furthermore, there could be cultural or religious reasons why participants are not willing to say a health state is worse than death. Further work is encouraged to investigate this further.

A final note regarding the practical importance of the differences in average health state utilities between the parametric and nonparametric models. The practical importance of these differences relies on their impact on the relative cost per QALY of multiple interventions. The impact on cost per QALY depends on the nature of the benefit, the severity of the patients' health state, cost, and how close the incremental cost per QALY of an intervention is to the willingness-to-pay threshold used by the funding body. Further research is then recommended to explore the consequences of the nonparametric Bayesian model for patient groups.[30]

Conclusion

The effort of data collection is the biggest cost of obtaining health state valuations, so it is a responsibility on the researcher to make the most of the data by fitting the best model. The nonparametric Bayesian model presented here provides better utility estimates over previously adopted standard parametric models. It is also more flexible in estimating the covariates' effect on health state values. In general, it provides a better approach to modeling health state values for other instruments such as EQ-5D and HUI-II, in addition to more specialized, condition-specific measures.

Article and Author Information

Accepted for Publication: February 19, 2021

Published Online: May 7, 2021

doi: https://doi.org/10.1016/j.vhri.2021.02.004

Author Affiliation: Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, Beirut, Lebanon (Kharroubi).

Correspondence: Samer A. Kharroubi, PhD, CStat, Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut. P.O.BOX: 11-0236, Riad El Solh 1107-2020, Beirut, Lebanon. Email: sk157@aub.edu.lb

Author Contributions: *Concept and design*: Kharroubi
Acquisition of data: Kharroubi
Analysis and interpretation of data: Kharroubi
Drafting of the manuscript: Kharroubi
Critical revision of the paper for important intellectual content: Kharroubi
Statistical analysis: Kharroubi
Provision of study materials or patients: Kharroubi
Obtaining funding: Kharroubi
Administrative, technical, or logistic support: Kharroubi
Supervision: Kharroubi

Conflict of Interest Disclosures: The author reported no conflict of interest.

Funding/Support: This work was supported by grant 103366 from the University Research Board at the American University of Beirut, Lebanon.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The author thanks the University Research Board, American University of Beirut for funding this study.

REFERENCES

1. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
2. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang QA. Multiattribute utility function for a comprehensive health status classification system: health Utilities Index Mark 2. *Med Care*. 1996;34(7):702–722.
3. Feeny DH, Furlong WJ, Torrance GW, et al. Multi-attribute and single-attribute utility function for the Health Utility Index Mark 3 system. *Med Care*. 2002;40(20):113–128.
4. Hawthorne G, Richardson G, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med*. 2001;33(5):358–370.
5. Kaplan RM, Anderson JP. A general health policy model: update and application. *Health Serv Res*. 1988;23(2):203–235.
6. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271–292.
7. Revicki DA, Leidy NK, Brennan-Diemer F, Sorenson S, Togias A. Integrating patients' preferences into health outcomes assessment: the multiattribute asthma symptom utility index. *Chest*. 1998;114(4):998–1007.
8. Brazier JE, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition specific measure: the King's Health Questionnaire. *Med Decis Mak*. 2008;28(1):113–126.
9. Dolan P. Modeling valuation for Euroqol health states. *Med Care*. 1997;35(11):1095–1108.
10. McCabe C, Stevens K, Roberts J, Brazier JE. Health state values for the HUI2 descriptive system: results from a UK Survey[published correction appears in *Health Econ*. 2006;15(7):761]. *Health Econ*. 2005;14(3):231–244.
11. Kharroubi SA, O'Hagan A, Brazier JE. Estimating Utilities from individual health state preference data: a nonparametric Bayesian approach. *Appl Stat*. 2005;54:879–895.
12. McGhee SM, J Brazier, CLK Lam, et al. Quality adjusted life years: population specific measurement of the quality component. *HK Med J*. 2011;17(suppl 6):17–21.
13. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–220.
14. Kharroubi SA, Brazier JE, Roberts J, O'Hagan A. Modelling SF-6D health state preference data using a nonparametric Bayesian method. *J Health Econ*. 2007;26(3):597–612.
15. Kharroubi SA, Brazier JE, O'Hagan A. Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. *Soc Sci Med*. 2007;64(6):1242–1252.
16. Kharroubi SA, McCabe C. Modelling HUI 2 health state preference data using a nonparametric Bayesian method. *Med Decis Mak*. 2008;28(6):875–887.
17. Kharroubi SA, Abou Daher C. Modelling a Preference-Based Index for EQ-5D using a non-parametric Bayesian method. *Qual Life Res*. 2018;27(11):2841–2850.
18. Brazier JE, Fukuhara S, Roberts J, Kharroubi SA, O'Hagan A, Ikeda S. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol*. 2009;62(12):1323–1331.
19. Kharroubi SA, Brazier JE, McGhee S. Modelling SF-6D Hong Kong standard gamble health state preference data using a nonparametric Bayesian method. *Value Health*. 2013;16(6):1032–1045.
20. Kharroubi SA, O'Hagan A, Brazier JE. A comparison of United States and United Kingdom EQ-5D health states valuations using a nonparametric Bayesian method. *Stat Med*. 2010;29(15):1622–1634.
21. Kharroubi SA, Brazier JE, McGhee S. A comparison of Hong Kong and United Kingdom SF-6D health states valuations using a nonparametric Bayesian method. *Value Health*. 2014;17(4):397–405.
22. Kharroubi SA. A comparison of Japan and United Kingdom SF-6D health states valuations using a nonparametric Bayesian method. *Appl Health Econ Health Policy*. 2015;13(4):409–420.
23. Kharroubi SA, Beyh Y, Harake E, Dawoud M, Rowen D, Brazier J. Examining the feasibility and acceptability of valuing the Arabic version of SF-6D in a Lebanese population. *Int J Environ Res Public Health*. 2020;17(3):1037.
24. Furlong W, Feeny D, Torrance GW, Barr R. *Horsman J Guide to Design and Development of Health State Utility Instrumentation. Centre for Health Economics and Policy Analysis Paper Series 1990-09*. Hamilton, Canada: McMaster University; 1992.
25. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. *Med Decis Mak*. 1994;14(1):9–18.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between measurement. *Biochim Clin*. 1987;11:399–404.
27. Lebanese Republic Ministry of Social Affairs/Central Administration of Statistics. Demographic and Social Statistics. http://www.cas.gov.lb/index.php/demographic-and-social-en; 2009. Accessed April 1, 2020.
28. Méndez I, Perpiñán JMA, Martínez FIS, Pérez JEM. Inverse probability weighted estimation of social tariffs: an illustration using the SF-6D value sets. *J Health Econ*. 2011;30(6):1280–1292.
29. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation and validation of the SF-36 Health Survey for use in Japan. *J Clin Epidemiol*. 1998;51(11):1037–1044.
30. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14(6):1523–1532.