VALUE IN HEALTH 22 (2019) 482–490

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Estimating Joint Health Condition Utility Values

Alexander J. Thompson, MSc [1,*], *Matthew Sutton, MSc, PhD* [2], *Katherine Payne, MSc, PhD* [1]

[1]Manchester Centre for Health Economics, Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Manchester, UK; [2]Health Organisation, Policy and Economics, Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Manchester, UK

A B S T R A C T

Objectives: To predict health state utility values (HSUVs) for individuals with up to 4 conditions simultaneously. **Methods:** Person-level data were taken from the General Practice Patient Survey, a national survey of adult patients registered with general practices in England. Individuals reported whether they had any 1 of 16 chronic conditions and completed the 3-level EuroQol 5-dimensional questionnaire. Four nonparametric methods (additive, multiplicative, minimum, and the adjusted decrement estimator) and 1 parametric estimator (the linear index) were used to predict HSUVs for individuals with a joint health condition (JHC). Predicted and actual utility scores were compared for precision using root mean square error and mean absolute error. Bias was assessed using mean error. **Results:** The analysis included 929,565 individuals, of which 30.5% had at least 2 conditions. Of the nonparametric estimators, the multiplicative approach produced estimates with the lowest bias and most precision for 2 JHCs. For populations with a long-term mental health condition within the JHC, the multiplicative approach overestimated utility scores. All nonparametric methods produced biased results when estimating HSUVs for 3 or 4 JHCs. The linear index generally produced unbiased results with the highest precision. **Conclusions:** The multiplicative approach was the best nonparametric estimator when estimating HSUVs for 2 JHCs. None of the nonparametric approaches for estimating HSUVs can be recommended with more than 2 JHCs. The linear index was found to have good predictive properties but needs external validation before being recommended for routine use.

Keywords: additive, EQ-5D, linear index, minimum, multimorbidity, multiple conditions, multiplicative, utility

Introduction

With an aging population, the prevalence of patients with multiple simultaneous chronic conditions (commonly called multimorbidity) is growing.[1] Decision-analytic models designed to assess the cost effectiveness of interventions for long-term chronic conditions need to explicitly quantify the impact of "joint health conditions" (JHCs) that patients who are multimorbid would transition between overtime within the model structure.[2] A key practical constraint for quantifying the impact of JHCs is the availability of health state utility values (HSUVs). It is not feasible to elicit all potential combinations of JHCs because of the potential for "combinatorial explosion."[3] Consequently, a number of nonparametric and parametric methods have been proposed to estimate HSUVs for JHCs using available data from populations with only single conditions.[4–6]

In 2013, a published systematic review summarized available approaches to estimate utility values for JHCs and identified only 11 relevant studies.[6] The review found 5 different methods for combining HSUVs for JHCs from single conditions. These include 4 nonparametric methods: additive (individual disutility values are summed), multiplicative (individual utility values are multiplied), minimum (lowest individual utility value is selected), and adjusted decrement estimator (ADE; combination of additive, multiplicative, and minimum and using an adjustment factor). In practice, the adjustment factor used in the ADE can also be calculated using regression-based parametric methods. The fifth method included in the review was a parametric approach (the linear index), which uses regression analysis to assign weights to 3 elements underpinned by the additive, multiplicative, and minimum methods.[3,7–15]

Source of financial support: No financial support was received for this study.

* *Address correspondence to*: Alexander J. Thompson, MSc, Manchester Centre for Health Economics, Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Room 4.313, Jean Mcfarlane Bldg, Oxford Rd, Manchester M13 9PL, UK.

E-mail: alexander.thompson@manchester.ac.uk

https://doi.org/10.1016/j.jval.2018.09.2843

The review observed that the additive and multiplicative methods were the most commonly assessed methods to estimate HSUVs for JHCs from populations with only single conditions. The studies included in the review presented contradictory findings, dependent on the method, which could lead to substantial bias in the results of decision-analytic models that use JHCs within their model structure. Moreover, the review found only 1 article that investigated estimating HSUVs for health states with more than 2 JHCs.[15] Because of the paucity of empirical research, caution has been urged when estimating HSUVs for populations with more than 2 JHCs.[4]

The primary aim of this study was to identify which of the available nonparametric and parametric methods within the current literature generated unbiased and precise estimates of HSUVs for individuals with up to 4 JHCs. The secondary aim was to assess whether the presence of particular conditions within the JHC had an impact on bias and precision.

Methods

All analyses were conducted in Stata version 14.1 (Stata Corp, College Station, TX).

Data

The analysis was conducted on a representative sample, the weighted responses, from patients in the General Practice Patient Survey (GPPS) for wave 2 in the year 2010 to 2011 (potential n = 1 037 946). A survey, designed to gather patients' views and experiences on using services of the local National Health Service including quality of care, access, and service planning, was posted to all patients registered with a general practice for more than 6 months in England. Patients were included in this analysis data set only if they had completed the 3-level EuroQol 5-dimensional questionnaire (EQ-5D-3L) and had a full set of condition characteristics. Missing data were not imputed. The total eligible sample size was 929 565 individuals.

Outcome Measure

The primary outcome measure for this analysis was a utility value generated from the EQ-5D-3L. Patients completing the GPPS were asked to rate their current health status using the EQ-5D-3L. The published UK national tariff for the EQ-5D-3L, which is based on public preferences of a sample of the UK population, was then applied to give utilities ranging from 1 (equivalent to "full health") to −0.594.[16] In the EQ-5D-3L, a health state equivalent to death is anchored at 0.

Health Conditions

The GPPS asks respondents whether they have any of 16 specific long-term health conditions. The analysis was then based on JHC predictions for subgroups of individuals with 2 conditions (120 potential unique JHC combinations), 3 conditions (560 potential unique JHC combinations), and 4 conditions (1820 potential unique JHC combinations). Some combinations of conditions were very rare even within this very large data set. Consequently, mean

Table 1 – Summary statistics of the analysis sample.

Variable	All patients	No conditions	1 condition	2 conditions	3 conditions	4 conditions
Observations, N (%)	929 565	408 103 (43.9)	273 700 (29.4)	136 559 (14.7)	65 247 (7.0)	28 656 (3.1)
Sex, male (%)	43.5	40.4	46.1	46.1	45.4	45.3
Age (y) (%)						
18-24	4.7	8.3	2.9	1.1	0.5	0.3
25-34	10.6	18.2	6.9	2.8	1.4	0.9
35-44	14.2	21.6	11.7	6.1	3.6	3.0
45-54	17.9	21.4	18.5	13.1	10.0	8.5
55-64	20.5	16.6	24.2	24.1	21.9	20.1
65-74	18.0	9.3	21.4	27.9	29.7	29.2
75-84	10.7	3.6	11.3	18.8	24.0	26.6
>85	3.4	0.9	3.0	6.0	8.9	11.3
EQ-5D-3L utility, mean ± SD	0.804 ± 0.265	0.922 ± 0.144	0.813 ± 0.227	0.693 ± 0.277	0.556 ± 0.319	0.431 ± 0.34
Long-term condition (%)						
Alzheimer disease	0.7		0.5	1.1	1.9	2.9
Angina or long-term heart problem	7.0		4.1	11.3	21.2	33.5
Arthritis or long-term joint problem	17.2		11.4	32.7	53.8	69.6
Asthma or long-term chest problem	10.8		12.5	17.4	24.0	32.5
Blindness or severe visual impairment	1.3		0.5	1.7	4.0	7.5
Cancer in the last 5 years	4.0		4.0	7.0	10.1	12.7
Deafness or severe hearing impairment	5.0		2.4	7.9	16.3	26.0
Diabetes	8.9		6.7	16.6	25.1	33.7
Epilepsy	1.1		1.1	1.8	2.6	3.5
High blood pressure	24.5		25.2	47.9	60.8	69.8
Kidney or liver disease	1.8		0.9	2.6	5.5	9.5
Learning difficulty	0.8		0.7	1.3	1.9	2.7
Long-term back problem	11.3		8.2	19.2	33.6	47.8
Long-term mental health problem	3.7		4.0	6.0	8.3	11.2
Long-term neurological problem	1.9		1.6	2.8	4.9	7.6
Another long-term condition	12.7		16.4	22.7	26.2	29.6

EQ-5D-3L indicates 3-level EuroQol 5-dimensional questionnaire.

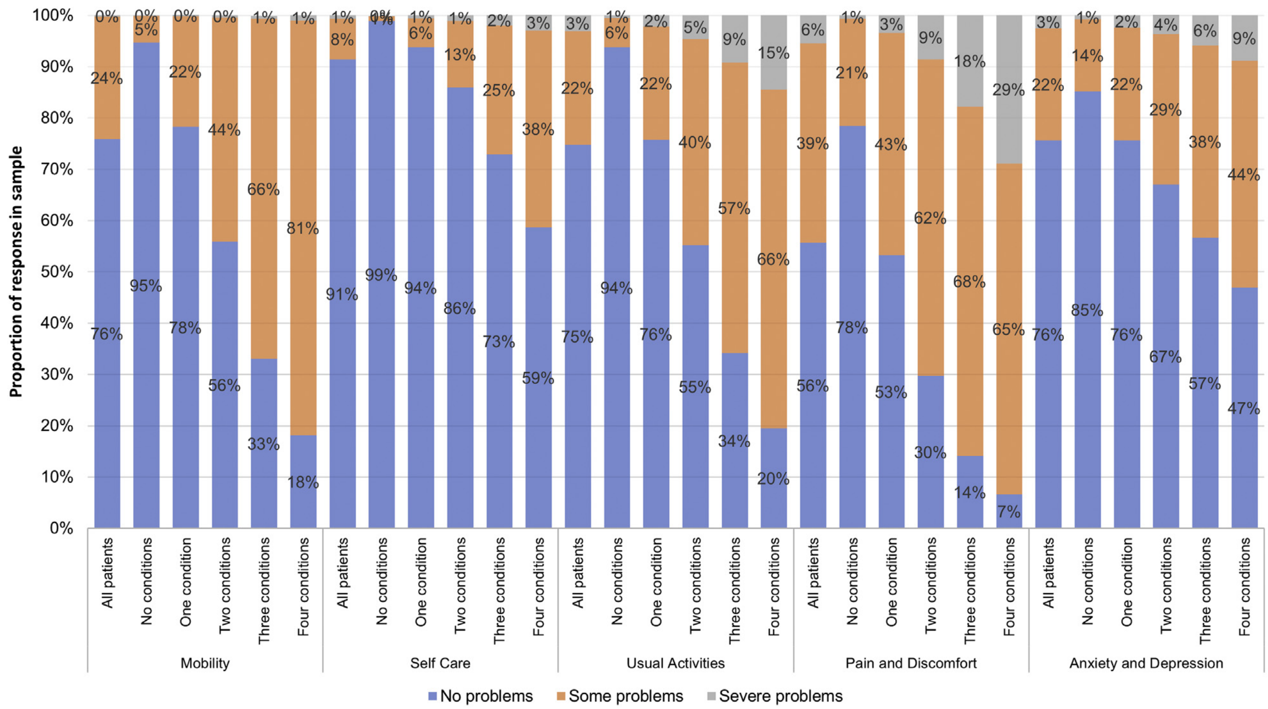

Fig. 1 – The impact of the number of health conditions on the EQ-5D-3L profile. EQ-5D-3L indicates 3-level EuroQol 5-dimensional questionnaire.

Table 2 – Base-case results: bias and precision for 2, 3, and 4 conditions.

Variable	2 conditions	3 conditions	4 conditions
	Mean ± SD (min. to max.)	Mean ± SD (min. to max.)	Mean ± SD (min. to max.)
Utility value	0.644 ± 0.12 (0.38 to 0.84)	0.548 ± 0.13 (0.19 to 0.78)	0.449 ± 0.15 (0.05 to 0.74)
Additive estimate	0.626 ± 0.14 (0.29 to 0.86)	0.523 ± 0.15 (0.06 to 0.80)	0.408 ± 0.18 (−0.19 to 0.73)
Multiplicative estimate	0.651 ± 0.12 (0.40 to 0.86)	0.579 ± 0.11 (0.30 to 0.81)	0.509 ± 0.11 (0.22 to 0.75)
Minimum estimate	0.720 ± 0.10 (0.57 to 0.89)	0.711 ± 0.08 (0.57 to 0.88)	0.702 ± 0.06 (0.58 to 0.87)
Adjusted decrement estimate	0.699 ± 0.11 (0.50 to 0.89)	0.639 ± 0.11 (0.44 to 0.00)	0.585 ± 0.10 (0.39 to 0.86)
Linear estimate	0.649 ± 0.12 (0.38 to 0.86)	0.554 ± 0.12 (0.23 to 0.80)	0.450 ± 0.13 (0.01 to 0.72)
RMSE			
Additive	0.050	0.067	0.085
Multiplicative	0.041	0.064	0.088
Minimum	0.096	0.186	0.276
ADE	0.080	0.151	0.219
Linear	0.042	0.056	0.063
MAE			
Additive	0.038	0.052	0.065
Multiplicative	0.031	0.050	0.071
Minimum	0.081	0.165	0.253
ADE	0.067	0.134	0.201
Linear	0.032	0.045	0.051
ME			
Additive	0.018	0.025	0.041
Multiplicative	−0.007	−0.032	−0.060
Minimum	−0.076	−0.163	−0.253
ADE	−0.061	−0.132	−0.201
Linear	−0.005	−0.006	0.000
Observations	106	165	108
Top ranked	Multiplicative	Linear	Linear

ADE indicates adjusted decrement estimator; MAE, mean absolute error; max., maximum; ME, mean error; min., minimum; RMSE, root mean square error.

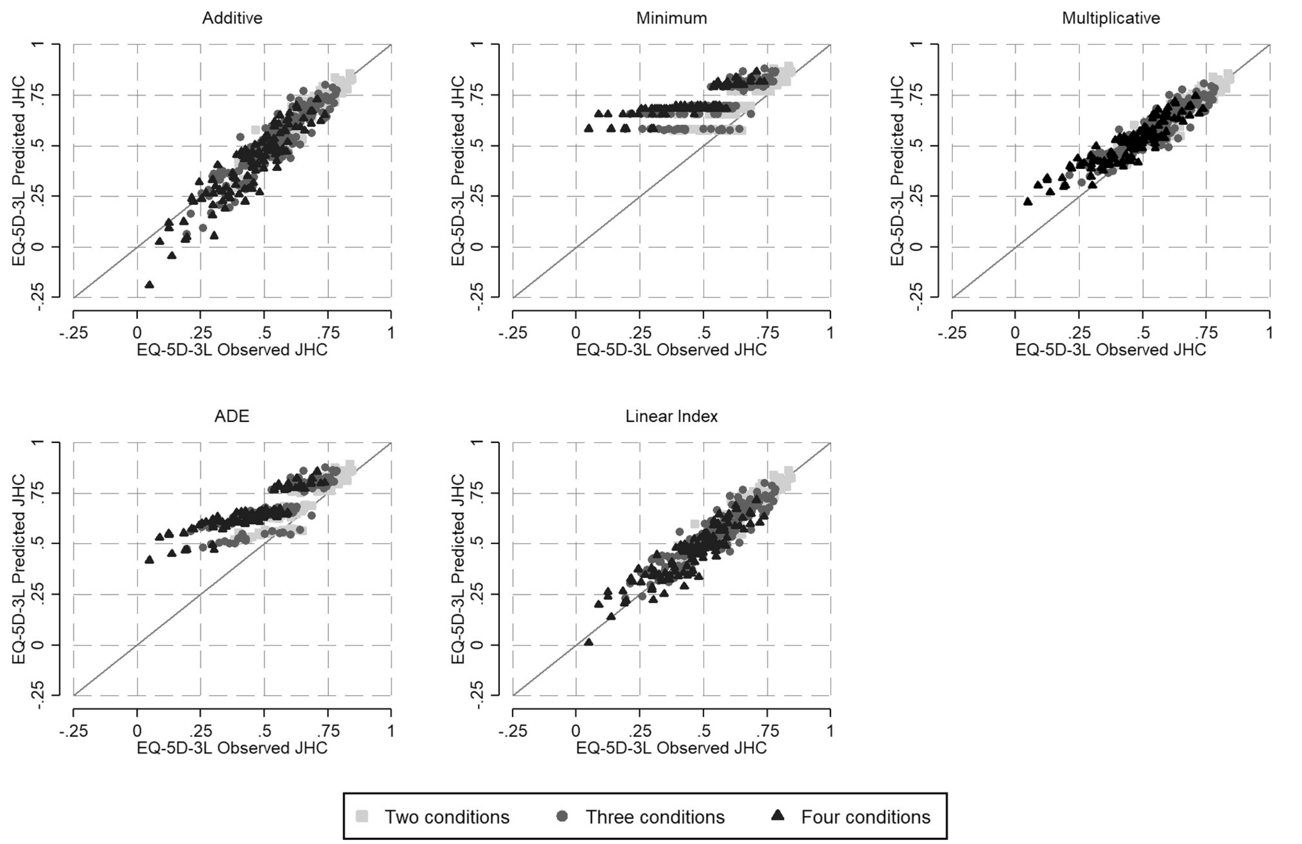

Fig. 2 – Predicted JHC for EQ-5D-3L utility versus observed JHC for EQ-5D-3L utility. Results are from condition combinations with at least 50 individuals. Rows represent the different methods, and columns represent the number of conditions. Estimates falling along the 45° line represent a perfect prediction. ADE indicates adjusted decrement estimator; EQ-5D-3L, 3-level EuroQol 5-dimensional questionnaire; JHC, joint health condition.

utility values were estimated for JHCs only where at least 50 patients had the JHC combination.

Estimating HSUVs: Using Nonparametric Approaches

The approach for calculating the nonparametric estimators involved identifying subgroups of individuals from within the entire data set, with and without the conditions of interest, using the equations from the different estimators (for formulas, see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.09.2843) to calculate the joint HSUVs and comparing the errors across the estimators. The entire data set was used for the nonparametric estimators because the simple mathematical approaches did not involve distributional assumptions or have the potential for "overfitting" the models. The following 6 steps were used:

1. The set of all potential combinations for individuals with a total morbidity count of up to 4 conditions was identified.
2. Individuals with all, and only, the conditions of interest were identified, starting with 2 conditions, 3 conditions, and then 4 conditions. Individuals with these combinations were identified and labeled and the total number of individuals was counted. The mean utility value for the combinations of conditions was estimated.
3. Individuals with only the single conditions of interest within the larger set of 2 conditions, 3 conditions, and 4 conditions were identified and labeled. The number of patients was counted, and the mean utility value was calculated.
4. The mean values from the single health conditions for the JHCs were estimated using different methods to predict the joint HSUVs.
5. Utility values for individuals with the observed JHCs and the estimates for the JHCs were compared.
6. Measures of error were computed across the whole data set between the predicted values and the observed values.

Estimation for the Parametric Estimator

The linear index uses linear regression to fit the additive, multiplicative, and minimum approaches to the underlying utility data and so can be considered a parametric approach.[14] The parameter weights are estimated on the study data set, which may mean that the weights lack generalizability when being applied to other data sets containing different populations, baseline utilities, and condition combinations. In this study, we used k-fold cross-validation to create k ($k = 10$) estimation and validation data sets. k-fold cross-validation involves splitting the whole data set into k approximately equal "folds," with the first fold being used for prediction and the remaining $k - 1$ folds being used to estimate the model.[17] Linear regression models, using ordinary least squares, were ran separately on 10 estimation data sets containing data on single and JHC utility values. The linear models were then used to predict utility values for the JHCs in the remaining $k - 1$ prediction sets. Measures of error were computed between the predicted utility values and the actual values in the prediction data set.

Table 3 – Sensitivity results: bias and precision for specific conditions.

Variable	Alzheimer disease	Angina	Arthritis	Asthma	Blindness	Cancer	Deafness	Diabetes	Epilepsy
All combination JHCs									
Utility value	0.502	0.575	0.445	0.548	0.569	0.635	0.596	0.569	0.651
Linear estimate	0.461	0.569	0.442	0.583	0.583	0.586	0.592	0.587	0.679
Multiplicative estimate	0.476	0.601	0.488	0.610	0.607	0.615	0.621	0.617	0.686
RMSE: linear	**0.063**	0.053	0.056	0.060	0.044	**0.065**	0.052	0.052	0.054
MAE: linear	0.049	0.041	0.045	0.048	0.035	**0.056**	0.040	0.042	0.044
ME: linear	**0.041**	0.006	0.002	−0.035	−0.015	**0.049**	0.004	−0.017	−0.028
RMSE: multiplicative	0.054	0.059	0.074	**0.085**	0.052	0.043	0.055	0.070	0.060
MAE: multiplicative	0.043	0.045	0.056	**0.070**	0.045	0.036	0.043	0.055	0.048
ME: multiplicative	0.026	0.026	0.044	**0.063**	0.038	0.020	0.025	0.048	0.035
2 JHCs									
Utility value	0.513	0.692	0.563	0.722	0.626	0.702	0.719	0.690	0.687
Linear estimate	0.485	0.691	0.567	0.734	0.662	0.676	0.723	0.714	0.720
Multiplicative estimate	0.493	0.690	0.571	0.734	0.660	0.676	0.723	0.713	0.718
RMSE: linear	0.051	0.036	0.030	0.045	0.046	0.036	0.034	0.037	**0.060**
MAE: linear	0.037	0.027	0.025	0.037	0.038	0.030	0.026	0.031	**0.046**
ME: linear	**0.028**	0.001	−0.004	−0.012	**−0.035**	0.026	−0.004	−0.024	**−0.032**
RMSE: multiplicative	**0.050**	0.035	0.029	0.044	0.045	0.034	0.033	0.036	**0.059**
MAE: multiplicative	0.038	0.026	0.023	0.036	0.038	0.030	0.025	0.030	**0.045**
ME: multiplicative	0.020	0.002	−0.008	−0.012	**−0.034**	0.026	−0.004	−0.023	**−0.031**
Observed, all (2 JHCs)	16 (10)	89 (14)	145 (15)	96 (15)	38 (11)	58 (14)	87 (15)	97 (15)	19 (12)

Note. Highlighted results in boldface represent the top 3 worst-performing results (highest error) across the conditions.
JHC indicates joint health condition; MAE, mean absolute error; ME, mean error; RMSE, root mean square error.

Picking the Optimal Prediction Method: Measure of Precision and Bias

The mean absolute error (MAE), root mean square error (RMSE), and mean error (ME) were calculated to assess predictive ability. Both RMSE and MAE are measures of accuracy or precision and represent the average prediction error in the unit of interest.[18] Both metrics range from 0 to infinity, with lower errors being preferred. RMSE gives a relatively high weight to larger errors that are further away from the true estimate, whereas MAE is less sensitive to these outliers. By comparison, ME is a measure of systematic estimation bias and can be either positive or negative, with the sign of ME indicating whether the estimators systematically underestimates or overestimates the true JHC utility values.[18]

A similar approach adopted by Davison et al[19] was used to rank the different estimators for precision. Each estimator's RMSE and MAE were ranked, with the lowest being preferred. A joint ranking based on these precision measures was produced by summing the individual rankings. The model with the best ranking was deemed to be optimal. MEs as well as plots of predicted versus observed were used to judge whether any of the estimators were biased.

For the linear index approach, calculations of RMSE, MAE, and ME were based on the MEs found within the *k*-fold prediction data sets. Rankings were based on mean prediction errors across the *k*-fold data sets.

Setting Baseline Utility

The calculations for both parametric and nonparametric estimators required an input for baseline utility for healthy patients. A range of different methods to proxy for patients without the condition of interest exist, including the use of general population values.[20–22] In this analysis, the mean utility value for individuals completing the GPPS who selected themselves as having none of the 16 conditions listed was used as the baseline utility value to represent a "healthy" patient.

Sensitivity Analysis: Assessing Model Fit for Individual Conditions

An additional analysis assessed whether the optimal model(s) from the primary analysis for 2 conditions, 3 conditions, and 4 conditions was robust to the presence of particular conditions within the JHCs. Measures of error (RMSE, MAE) and precision (ME) were stratified for the optimal model for each of the 16 conditions within the data set. Observed versus predicted plots were also calculated for each of the 16 conditions for the optimal model(s).

Results

Descriptive Analysis

Table 1 presents an overview of the characteristics of the included patients (n = 929 565). There was a greater proportion of females (56.5%) than males (43.5%), and individuals were most likely to be in the age group of 55 to 64 years (20.5%). A total of 521 462 (56.1%) individuals reported as having at least 1 chronic condition, with 283 762 (30.5%) individuals having 2 or more chronic conditions. The most commonly reported condition was high blood pressure (see Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.09.2843), followed by arthritis or a long-term joint problem. For patients with 4 conditions, 69.8% and 69.6% had high blood pressure and arthritis, respectively.

EQ-5D-3L Profile and Utility Values

The mean utility score generated from the EQ-5D-3L profiles was 0.804 ± 0.265. For patients without a long-term condition, the

High blood pressure	Kidney or liver disease	Learning difficulty	Long-term back problem	Long-term mental health problem	Long-term neurological problem	Another long-term condition
0.579	0.594	0.590	0.428	0.380	0.411	0.507
0.568	0.593	0.605	0.437	0.436	0.386	0.519
0.600	0.618	0.616	0.485	0.472	0.426	0.556
0.052	0.042	0.046	0.061	**0.078**	0.051	0.055
0.042	0.038	0.033	**0.049**	**0.066**	0.042	0.044
0.011	0.002	−0.015	−0.009	**−0.055**	0.025	−0.012
0.059	0.047	0.060	**0.086**	**0.115**	0.067	0.077
0.045	0.037	0.044	**0.065**	**0.099**	0.053	0.059
0.021	0.024	0.026	**0.056**	**−0.092**	0.016	0.049
0.751	0.683	0.659	0.574	0.519	0.519	0.649
0.736	0.700	0.659	0.579	0.545	0.497	0.660
0.737	0.700	0.661	0.582	0.549	0.503	0.662
0.030	0.038	0.040	0.040	**0.052**	**0.051**	0.036
0.022	0.033	0.027	0.031	**0.041**	**0.046**	0.026
0.015	−0.017	0.000	−0.006	−0.026	0.022	−0.011
0.029	0.037	0.040	0.037	**0.051**	0.048	0.034
0.021	0.033	0.028	0.028	**0.040**	**0.041**	0.026
0.014	−0.017	−0.002	−0.008	**−0.030**	0.016	−0.012
157 (15)	36 (10)	13 (9)	105 (15)	54 (15)	32 (12)	97 (15)

mean utility was 0.922 ± 0.144. Utility values decreased in an ordinal fashion, as expected, with an increasing number of JHCs.

Figure 1 shows the impact of the number of conditions on the domains within the EQ-5D-3L profile. Across the whole sample, the pain/discomfort domain was the most likely to be causing individuals "some problems" or "extreme problems." The pain/discomfort as well as the mobility domains were the most sensitive to change depending on the number of conditions reported, with only 20% and 18% of respondents, respectively, reporting no problems in these domains when 4 conditions were reported. The least responsive domain to an increase in the number of conditions was the self-care domain, with 59% of individuals reporting no problems even when reporting 4 long-term conditions.

Base-Case Analysis: Predicting JHC Utility Values

Table 2 presents the results for JHCs with at least 50 patients in the sample. For 2 conditions, 106 of the possible 120 JHCs had at least 50 individuals. The most common co-occurring conditions were arthritis and high blood pressure (n = 15 266). Overall, the minimum method tended to estimate the highest JHC utility values for 2 conditions, whereas the additive method tended to estimate the lowest JHC utility values. The multiplicative performed best when ranking both the RMSE and MAE for precision for 2 conditions. The linear index did best when considering ME.

For 3 conditions, 165 of the possible 520 JHCs were recorded by at least 50 individuals. The most commonly co-occurring conditions were arthritis, high blood pressure, and back pain (n = 5027). The linear index performed best for precision on both the RMSE and the MAE. The low ME estimate indicated no sign of bias for the linear estimator. The multiplicative approach was the second-best option on the basis of precision, but the ME statistic demonstrated some evidence of overestimated scores. The plot for the multiplicative estimator for 3 conditions showed that the bias was more likely when the JHC had a lower expected utility value. The minimum and ADE approaches did worst in the measures of precision and demonstrated evidence of bias, systematically overestimating the utility values for JHCs. The additive approach had a middling ranking for precision. Nevertheless, for 3 conditions the additive approach was beginning to demonstrate evidence of systematically underestimating utility values for JHCs, particularly when the JHC utility value was expected to be at the lower end of the distribution.

For 4 conditions, only 108 of the possible 1820 JHCs were recorded by at least 50 individuals. The most common JHCs were arthritis, diabetes, high blood pressure, and back pain (n = 1237). The linear index performed best for precision and showed no sign of bias for 4 conditions. The multiplicative approach was the second-best option on the basis of precision (Table 2) but demonstrated strong evidence of bias (Fig. 2). Utility values were overestimated particularly when the utility values were at the lower end of the distribution. The minimum and ADE approaches did worst for the measures of precision and these approaches demonstrated strong evidence of systemically overestimating the utility values for JHCs. The additive approach again had a middling ranking for precision and demonstrated strong evidence of systematically underestimating utility values for JHCs, particularly when the JHC utility value was at the lower end of the distribution.

Sensitivity Analysis: Model Fit for Particular Conditions Within the Set of JHCs

Table 3 presents results for the linear index and the multiplicative estimator (the 2 best estimators from the base-case analysis) stratified by the presence of particular conditions within the JHC. Both estimators perform less well when stratified by condition than when applied to all JHCs. When considering 2 conditions, the ME for both the multiplicative estimator and the linear index was equal to or exceeded (absolute) 0.015 for Alzheimer disease,

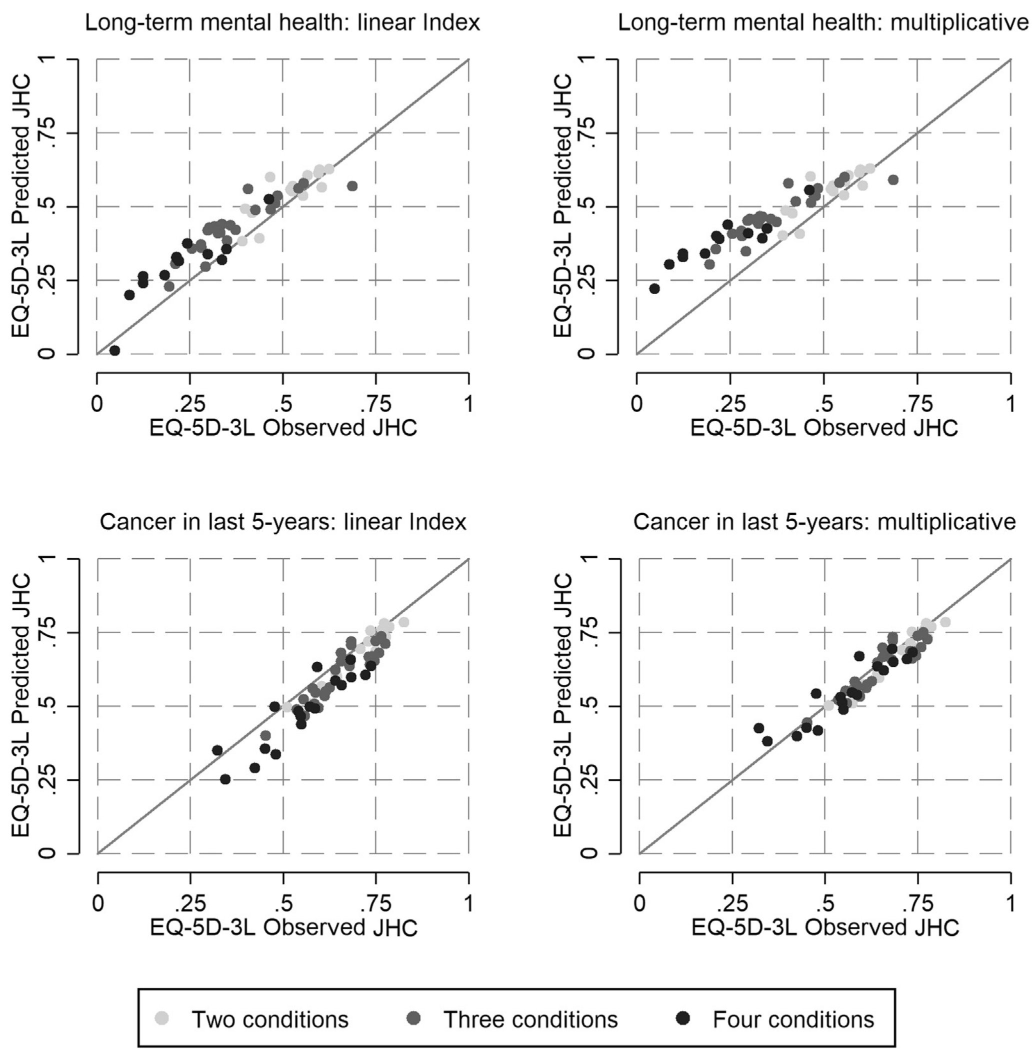

Fig. 3 – Predicted JHC for EQ-5D-3L utility versus observed JHC for EQ-5D-3L utility for patients with a long-term mental health condition and cancer in last 5 y falling within the set of JHCs. Results are from condition combinations with at least 50 individuals. Estimates falling along the 45° line represent a perfect prediction. EQ-5D-3L indicates 3-level EuroQol 5-dimensional questionnaire; JHC, joint health condition.

cancer, blindness, diabetes, epilepsy, high blood pressure, kidney or liver disease, long-term mental health problem, and long-term neurological problem. The ME was also equal to 0.015 for high blood pressure for the linear index. Such an error was sufficient to reject the additive approach for use in populations with 2 conditions in the base-case analysis because of systematic bias.

The highlighted outputs from Table 3 demonstrate that for patients with a long-term mental health condition, both the multiplicative estimator and the linear index tended to produce estimates that were in the top 3 worst for both bias and precision across conditions. Figure 3 and Table 3 show that for patients with a long-term mental health condition, the linear index is preferred to the multiplicative estimator but both estimators tend to overpredict utility, and the performance of the linear index and the multiplicative estimator gets worse depending on the number of conditions within the JHC. In comparison, for patients with cancer, the multiplicative estimator does better, but both estimators tend to underpredict utility. Although the performance of the linear index gets worse depending on the number of conditions, the multiplicative estimator seems to perform equally well irrespective of the number of other conditions within the JHC.

Discussion

This study has made use of survey data containing nearly 1 million individuals from the general population in England. The available sample size facilitated a robust investigation into the accuracy of 4 nonparametric approaches (additive, multiplicative, minimum, and ADE) and 1 parametric approach (linear index) commonly used within the literature for estimating joint HSUVs. Previous studies have largely been conducted on much smaller data sets, with analysis based on cohorts containing hundreds,[3,14] thousands,[7] or tens of thousands[8–13] of individuals. The available data for this study facilitated investigations of up to 4 conditions simultaneously, which has been suggested as an important area for research.[4]

Methods to combine HSUVs need to produce nonbiased estimates, also ideally with high precision. This study finds that of the simple nonparametric approaches, the multiplicative method performed best for individuals with 2 conditions simultaneously. There is a small risk of bias, particularly for JHCs that cause greater disutility when the multiplicative approach may underestimate total disutility. For precision, the multiplicative approach was the most precise estimator from the 5 estimation approaches. These findings support previous recommendations for the use of the multiplicative approach to estimate HSUVs for 2 conditions simultaneously.[4–6] Beyond 2 conditions, the results suggested that none of the simple nonparametric approaches can be recommended because each approach clearly leads to biased estimates, particularly at the lower end of the EQ-5D-3L utility distribution.

When analyzing the results by condition counts, the overall patterns of error for higher numbers of conditions are as expected. The additive approach tends to lead to underestimates of utility, likely reflecting a breakdown in the structural assumption that each individual condition acts independently on disutility. For higher numbers of conditions, some overlap between the underlying attributes underpinning the conditions is highly likely and so the independence assumption breaks down. In comparison, the minimum approach (and the ADE approach that places significant importance on the minimum) tends to lead to overestimates of utility as it becomes increasingly unlikely that 1 condition, as the minimum approach assumes, can be representative of all the negative attributes associated with a set of conditions.

The results from the sensitivity analysis suggested that researchers should exercise caution when combining HSUVs for certain conditions, even when using the optimal estimators identified in this article. In particular, when long-term mental health was included within a JHC, neither the linear index nor the multiplicative approach produced unbiased estimates, even for 2 JHCs. In general, the multiplicative approach provides overestimates of utility, making populations with a long-term mental health condition, alongside another condition, appear healthier than they actually are. This overestimation could have important consequences for studies estimating the burden of disease[23] or studies that aim to assess the cost effectiveness of interventions treating patients with depression.

The reason for the poorer performance of the estimators in individuals with a mental health condition is unknown. One potential reason could be due to the construction of the EQ-5D-3L, with its depression/anxiety domain being sensitive to individuals with a long-term mental health condition, alongside another condition. In comparison, no other condition within the GPPS has a directly matching domain within the EQ-5D-3L. Alternatively, there could be a genuine “synergistic” interaction effect,[6] whereby the total impact of depression on HSUVs, when experienced with other conditions, is greater than the sum of the independent condition effects. It is known that depression can exacerbate the pain associated with a chronic physical health problem and has been found to lead to poorer mortality outcomes for patients with cardiovascular disease. The National Institute for Health and Care Excellence has consequently developed a clinical guideline on depression with a chronic physical health problem to manage this population.[24]

The analysis in this study was conducted on a large sample from an English population, but there are some limitations that may reduce generalizability. Uncertainty in JHC estimates can originate not only from the method used to combine single health states but also from the system used to describe health as well as the preference weights or value set attached to the descriptive system. Therefore, the results from this study may not generalize to other utility measures with different descriptions of health such as the 5-level EQ-5D (EQ-5D-5L), 6-dimensional health state short form (SF-6D), and health utilities index 3 (HUI3). Moreover, even when the same descriptive system has been used, generalizability is likely to be limited when a different set of preference weights has been used, such as the US value set for the EQ-5D-3L. It is also feasible that populations from different countries may systematically complete health questionnaires in different ways.

The results showing the linear index as being optimal for higher numbers of conditions were based on *k*-fold cross-validation within the same data set. Before the linear index can be recommended, the coefficients (published in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.09.2843) should be validated in a separate data set to assess external validity for 2 conditions, 3 conditions, and 4 conditions.

The analysis in this article was based on self-reported conditions from a patient survey rather than on medical claims, primary care records, or interviews. The use of self-report can lead to a number of potential limitations and/or problems of generalizability. First, there was no information on disease severity or grading within the survey, which can have an important impact on disutility. Second, there are known differences between estimates of prevalence of morbidity based on self-report studies versus medical records[25,26] and patient and clinician interviews.[27] Such findings suggest the potential for overdiagnosis or underdiagnosis, which in turn can have an impact on utility estimates for JHCs, particularly if underdiagnosis or overdiagnosis is associated with levels of morbidity. Third, the disutilities associated with increasing counts of morbidity in this study are much larger than those found in other studies in which patient records and the *International Classification of Diseases* diagnosis codes have been used.[12,28,29] Consequently, it is not certain that findings here will generalize to utility values based on “catalogs” linked to diagnosis codes.[28,30]

Finally, this study investigated combinations with up to 4 conditions simultaneously. It is not certain that the trends found in this study, worsening performance in higher JHCs, would continue into even higher numbers of JHCs although it would seem reasonable to assume so.

Conclusions

This study has compared the performance of all nonparametric and parametric approaches proposed in the literature to estimate JHC utility values from single-condition populations. The study adds to the current literature by making use of an extremely large data set to assess combinations of up to 4 conditions and by assessing predictive accuracy stratified by condition type. Overall, the multiplicative approach was the best nonparametric approach for predicting JHCs for 2 conditions. Nevertheless, for populations with a long-term mental health condition, the multiplicative approach is likely to overestimate utility. For populations with 3 and 4 conditions, none of the nonparametric approaches can be recommended because there was clear bias in the predictions. The linear index had the best predictive properties overall but still produced biased estimates for particular conditions, notably long-term mental health. The linear index cannot be recommended for higher numbers of conditions before validation on a different data set has been conducted.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.09.2843.

REFERENCES

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012;380(9836):37–43.
2. Guthrie B, Thompson A, Dumbreck S, et al. Better guidelines for better care: accounting for multimorbidity in clinical guidelines—structured

examination of exemplar guidelines and health economic modelling. Health Serv Deliv Res 2017;5(16):1–150.
3. Basu A, Dale W, Elstein A, Meltzer D. A linear index for predicting joint health-states utilities from single health-states utilities. Health Econ 2009;18(4):403–19.
4. Ara R, Brazier J. Estimating health state utility values for comorbidities. Pharmacoeconomics 2017;35(S1):89–94.
5. Ara R, Wailoo A. NICE DSU Support Document 12: The Use of Health State Utility Values in Decision Models. London: National Institute for Health and Care Excellence; 2011.
6. Ara R, Wailoo AJ. Estimating health state utility values for joint health conditions: a conceptual review and critique of the current evidence. Med Decis Making 2013;33(2):139–53.
7. Haagsma JA, Van Beeck EF, Polinder S, Toet H, Panneman M, Bonsel GJ. The effect of comorbidity on health-related quality of life for injury patients in the first year following injury: comparison of three comorbidity adjustment approaches. Popul Health Metr 2011;9:10.
8. Ara R, Brazier J. Estimating health state utility values for comorbid health conditions using SF-6D data. Value Health 2011;14(5):740–5.
9. Fu AZ, Kattan MW. Utilities should not be multiplied: evidence from the preference-based scores in the United States. Med Care 2008;46(9):984–90.
10. Hu B, Fu AZ. Predicting utility for joint health states: a general framework and a new nonparametric estimator. Med Decis Making 2010;30(5):E29–39.
11. Ara R, Brazier J. Comparing EQ-5D scores for comorbid health conditions estimated using 5 different methods. Med Care 2012;50(5):452–9.
12. Barra M, Augestad LA, Whitehurst DGT, Rand-Hendriksen K. Examining the relationship between health-related quality of life and increasing numbers of diagnoses. Qual Life Res 2015;24(12):2823–32.
13. Hanmer J, Vanness D, Gangnon R, Palta M, Fryback DG. Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. J Clin Epidemiol 2010;63(3):331–41.
14. Dale W, Basu A, Elstein A, Meltzer D. Predicting utility ratings for joint health states from single health states in prostate cancer: empirical testing of 3 alternative theories. Med Decis Making 2008;28(1):102–12.
15. Flanagan W, Mcintosh CN, Le Petit C, Berthelot J-M. Deriving utility scores for co-morbid conditions: a test of the multiplicative model for combining individual condition scores. Popul Health Metr 2006;4:13.
16. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: Results from a UK general population survey. 1995. Centre for Health Economics, University of York working paper 138. https://www.york.ac.uk/che/pdf/DP138.pdf. [Accessed February 1, 2018].
17. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer; 2009.
18. Walther BA, Moore JL, Walther BA. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/References/WaltherAndMoore2005.pdf.
19. Davison NJ, Thompson AJ, Turner AJ, et al. Generating EQ-5D-3L utility scores from the dermatology life quality index: a mapping study in patients with psoriasis. Value Health 2018;21(8):1010–8.
20. Ara R, Brazier J, Zouraq IA. The use of health state utility values in decision models. Pharmacoeconomics 2017;35:77–88.
21. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health 2010;13(5):509–18.
22. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value Health 2011;14(4):539–45.
23. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(380):2197–223.
24. Pilling S, Anderson I, Goldberg D, Meader N, Taylor C. Two Guideline Development Groups. Depression in adults, including those with a chronic physical health problem: summary of NICE guidance. BMJ 2009;339:b4108.
25. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. BMC Public Health 2013;13(1):251.
26. Barber J, Muller S, Whitehurst T, Hay E. Measuring morbidity: self-report or health care records? Fam Pract 2010;27:25–30.
27. Hansen H, Schäfer I, Schön G, et al. Agreement between self-reported and general practitioner-reported chronic conditions among multimorbid patients in primary care—Results of the MultiCare Cohort Study. BMC Fam Pract 2014;15(1):39.
28. Sullivan PW, Lawrence WF, Ghushchyan V. National catalog of scores in the United States for chronic conditions. Med Care 2005;43(7):736–49.
29. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006;26(4):410–20.
30. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making 2011;31(6):800–4.