Mapping Utility Scores from a Disease-Specific Quality-of-Life Measure in Bariatric Surgery Patients

Stefan Sauerland, MD, MPH,1 Sylvia Weiner, MD,2 Karin Dolezalova, MD,3 Luigi Angrisani, MD,4 Carlos Masdevall Noguera, MD,5 Manuel Garcia-Caballero, MD,6 Frédéric Rupprecht, MBA,7 Marc Immenroth, PhD7

1Institut für Forschung in der Operativen Medizin, University of Witten/Herdecke, Cologne, Germany; 2Abteilung für Chirurgie, Krankenhaus Sachsenhausen, Frankfurt, Germany; 3Centrum mininvasive chirurgie, Klinické centrum Iscare, Prague, Czechia; 4Chirurgia Generale ed Endoscopica, Ospedale S. Giovanni Bosco, Naples, Italy; 5Departamento de Cirugía, Hospital Universitario de Bellvitge, Barcelona, Spain; 6Departamento de Cirugía, Hospital Universitario de Málaga, Málaga, Spain; 7Ethicon Endo-Surgery (Europe) GmbH, Nordrhein-Westfalen, Germany

ABSTRACT

Objectives: To develop algorithms for a conversion of disease-specific quality-of-life into health state values for morbidly obese patients before or after bariatric surgery.

Methods: A total of 893 patients were enrolled in a prospective cross-sectional multicenter study. In addition to demographic and clinical data, health-related quality-of-life (HRQoL) data were collected using the disease-specific Moorehead-Ardelt II questionnaire (MA-II) and two generic questionnaires, the EuroQol-5D (EQ-5D) and the Short Form-6D (SF-6D). Multiple regression models were constructed to predict EQ-5D- and SF-6D-based utility values from MA-II scores and additional demographic variables.

Results: The mean body mass index was 39.4, and 591 patients (66%) had already undergone surgery. The average EQ-5D and SF-6D scores were 0.830 and 0.699. The MA-II was correlated to both utility measures (Spearman’s r = 0.677 and 0.741). Goodness-of-fit was highest (R² = 0.55 in the validation sample) for the following item-based transformation algorithm: utility (MA-II-based) = 0.4293 + (0.0336 × MA1) + (0.0071 × MA2) + (0.0053 × MA3) + (0.0107 × MA4) + (0.0001 × MA5). This EQ-5D-based mapping algorithm outperformed a similar SF-6D-based algorithm in terms of mean absolute percentage error (P = 0.045).

Conclusions: Because the mapping algorithm estimated utilities with only minor errors, it appears to be a valid method for calculating health state values in cost-utility analyses. The algorithm will help to define the role of bariatric surgery in morbid obesity.

Keywords: bariatric surgery, economics, EQ-5D, health status indicators, quality of life.

Introduction

Obesity is an increasingly common disorder in Western societies [1,2]. Health spending for overweight and obese people has risen rapidly, mainly because of the various comorbidities associated with obesity, such as type 2 diabetes, hypertension, dyslipidemia, gallstone disease, knee osteoarthritis, sleep apnea or even cancer [3]. Different degrees of obesity can be defined by the body mass index (BMI, defined as weight in kg/height in m²). Conservative measures, such as diets, are effective only for milder degrees of overweight and obesity [4,5]. Since over 15 years, there is consensus that morbid obesity (BMI > 40 kg/m² or BMI > 35 kg/m² with obesity-associated comorbidity) represents an indication for a weight-loss (bariatric) surgical procedure [5-8], such as gastric banding or gastric bypass. Today, bariatric surgery is a well-studied and increasingly popular treatment modality [9]. Between 1998 and 2002, the number of bariatric operations performed in the USA has risen nearly fivefold [10], and in Europe a similar rise can be expected from epidemiological data [11]. In many health-care systems, however, uncertainty about the cost-effectiveness and utility of bariatric surgery remained. In addition, it remains to be defined which type of bariatric procedure is best for which grade of obesity [9].

Because obesity affects nearly every aspect of our physical, mental, social, and emotional health, accurate assessment of quality-of-life is important when describing the effectiveness of weight-loss therapy [12,13]. The majority of literature on bariatric surgery assessed and reported health-related quality-of-life (HRQoL) using the Moorehead-Ardelt II questionnaire (MA-II), which is applicable to patients before or after weight-loss therapy [14,15]. The MA-II includes only six items and takes about a minute to complete. MA-II data can be combined with a score based on weight loss and complications.
and Reporting Outcome System (BAROS) can be used to define failure or success of surgery [16]. Most surgeons prefer the MA-II or other similar instruments over generic questionnaires, because disease-specific instruments offer a higher responsiveness [17]. The BAROS has been endorsed by various professional societies for use in clinical practice [18]. Currently, however, it is not possible to generate utilities from trials that have used the MA-II. Because utilities are the basis for the calculation of quality-adjusted life-years, they also are an essential part of cost-effectiveness analyses, health technology assessment, and in any health-care decision-making process.

The primary goal of this study therefore was to map the relationship between the MA-II and EQ-5D-based utility values. A calculation of health state values based on a conversion algorithm would allow health economists to use MA-II data from previous and future trials. A secondary goal was to compare the resulting algorithms between EQ-5D and SF-6D.

Methods

Study Design

The study was a cross-sectional survey conducted in four European countries between August and December 2007. Five leading centers of bariatric surgery took part in the study. The choice of centers was made taking into account the need to select from a variety of bariatric surgical procedures. By including patients with different cultural backgrounds, the study aimed at achieving better generalizability of results. Because a formal power calculation is not possible for a regression analysis, the sample size was decided to be approximately 800 patients. This number was chosen on practical grounds and on the basis of previous similar studies.

All patients were informed about the study by their surgeon and provided written informed consent to participate in the study. The study was approved by the Ethics Committee of the University of Witten/Herdecke. To ensure accuracy of medical data, the sites were visited during the study.

Study Sample

Because bariatric surgery can also lead to specific side effects impacting on HRQoL, the study sample included both preoperative and postoperative patients. This way of sampling allowed for a diversity of different health states and different degrees of obesity, even including patients who had returned to normal body weight. Patients were considered eligible if they either were scheduled for bariatric surgery or had received any bariatric operation in the past. All patients were morbidly obese, either currently or in the past. Nevertheless, we excluded patients who had had their index operation within the last 3 months, because the short-term sequelae of surgery might have interfered with accurate assessment of HRQoL. Each center consecutively included patients who were seen in the outpatient clinic.

Data were collected on sociodemographic (sex and age) and clinical characteristics (current weight, height, and metabolic, pulmonary, cardiovascular or other comorbidities). The severity of each comorbidity was defined as moderate (requiring no or only occasional medical therapy) or severe (requiring regular drug treatment). In postoperative patients, the preoperative status (weight and comorbidities) was recorded in addition to surgery-related data (date and type of surgery, complications). For those patients, who underwent a staged procedure (e.g., sleeve gastrectomy followed by biliopancreatic diversion with duodenal switch) or who had to be converted to another type of procedure, only the first operation was counted as index operation.

Questionnaires

The MA-II contains six questions, which were scored on a scale from 1 to 10. These questions address mood, physical function, social relationships, ability to work, sexuality, and eating behavior. In each question, the ends of the scale are labelled by descriptors and highlighted by colored icons. Higher scores represent better HRQoL. A summary score can be calculated, which ranges between 6 and 60 (or alternatively between −3 and 3). Values greater than 42 (or 1, respectively) indicate good HRQoL. Although bariatric surgeons in many countries currently are using the MA-II, so far it is validated only in English [14]. Therefore, translations of the MA-II into German, Czech, Italian, and Spanish were carried out with the permission of the copyright owners. In each country, these translations were carefully checked for accurate wording by the respective surgeons. The data generated in the present study will also be used to formally validate the translated versions of the MA-II. This validation analysis will follow in a subsequent publication. Given the simplicity of wording and the partly graphical format of the MA-II, however, it appears unlikely that the different language versions produced different results.

The EuroQoL-5D (EQ-5D) is a commonly used generic instrument which contains five items and a visual analogue scale (EQ-VAS) [19]. The EQ-5D defines five dimensions of health: mobility, self-care, usual activities, pain, and anxiety of depression. We used the weights published by Greiner et al. [20] to calculate a single summary score from the five descriptive items of the EQ-5D.

The Short Form-6D (SF-6D) is a preference-based measure of health, based on a subset of 11 questions from the SF-36. In the present study, country-specific versions of the revised instrument (SF-36v2 with 4-week recall) were used. Although the SF-36 measures eight dimensions of HRQoL [21], only six of
these dimensions form the basis of the SF-6D: by using the algorithm developed by Brazier et al. [22], we computed health state values from 11 items of the SF-36.

Missing data were replaced for the SF-36 by using the half-scale rule. When calculating the summary score for the MA-II, missing values for item 5 (sexuality) were estimated by replacing them with the average of the other five items. No other missing data were generated. The completeness of answers to the different questions exceeded 96% in all questionnaires.

**Model Selection**

Initially, the total sample was randomly subdivided into two equal sized groups. Regression models were developed on the first half of the sample, whereas the second half was used as cross-validation sample. Multivariate linear regression models were used to establish prediction models of EQ-5D and SF-6D. Because missing data were not estimated, the actual number of valuations varied in the different analyses. Utility index values were treated as continuous variables. To include HRQoL in the prediction model, the MA-II was used in two different ways: first, as a summary score, and second, as a set of six single items. Because we aimed for a simple-to-use and robust algorithm, only very few demographic variables were included in the models: age, sex, and current BMI. Additionally, the country (transformed into dummy variables) was studied. Furthermore, first-degree interaction terms were added to the model. To assess the influence of ceiling effects, we compared the results of regression analyses among pre- and postoperative patients, because postoperative patients are generally more likely to report HRQoL to be normal or even supernormal. Although HRQoL varies by social class and income, we chose not to incorporate such variables, because they are difficult to assess accurately in clinical routine.

The $R^2$ statistic provided a measure of the goodness-of-fit of the regression model. Regression residuals were examined to identify potentially suitable non-linear models. The absolute deviation between actual and predicted utility values served as a secondary indicator of model performance. Predicted values of EQ-5D exceeding 1.0 were truncated to this boundary value. The differences between observed and predicted utility values were divided by the observed value, thus calculating the mean (and median) absolute percentage error (MAPE). The MAPE was used as additional methods to describe the relative overall fit of the model.

**Additional Statistical Tests**

Comparisons between different subgroups of patients were carried out with Student’s $t$ and chi-square tests. The differences between EQ5D and SF6D validation results were analyzed by Wilcoxon signed-rank test for paired data. Correlations were evaluated by Pearson’s and Spearman’s coefficient of correlation. All calculations of $R^2$ were carried out using Pearson’s coefficient of correlation. All statistical analyses were performed using SPSS, version 15.0 (SPSS Inc., Chicago, IL, USA).

**Results**

**Patients**

Similar to other studies on morbid obesity, this cohort had had a clear preponderance of females (79%). The patients’ current BMI ranged from 17 to 75 kg/m², although all patients were obese before surgery. Further demographic information is presented in Table 1. As expected, both the BMI and the obesity-related comorbidities were significantly different between pre- and postoperative patients.

In the 592 postoperative patients, the BMI before surgery was 47.7 (SD 8.7). In these patients, a median of 12 months (range 3 to 196) had elapsed between the initial operation and the survey. There was a mixture of different bariatric procedures: gastric bypass (48.2%), gastric banding (29.8%), one-anastomosis gastric bypass (9.6%), sleeve gastrectomy (5.2%), biliopancreatic diversion (4.2%), and combined or other procedures (2.8%).

Health-related quality-of-life was significantly lower in patients before surgery than after surgery (Table 2). Both in pre- and postoperative patients, utilities were significantly lower when measured by SF-6D compared to EQ-5D. Although utilities using one of the two methods were closely correlated (Pearson’s $r = 0.682$, Spearman’s $r = 0.731$), the SF-6D gave higher utility scores than the EQ-5D. As shown in Figure 1, the difference ranged between 0.13 and 0.16, without a detectable influence of weight.

**Mapping the EQ-5D**

The EQ-5D was well correlated to the MA-II scale (Pearson’s $r = 0.662$, Spearman’s $r = 0.677$). The different regression models are presented in Table 3. The inclusion of additional covariates or nonlinear terms failed to improve the fit of these models. Specifically, the results of validation did not differ between pre- and postoperative patients, patients with or without comorbidity, or between different centers.

The $R^2$ value was largest (0.441), when the model was based on the six single items of the MA-II. Nevertheless, not all items contributed significantly to the fit of the model. To use the same variables as in the SF-6D-based model, the final model (Table 4) included only the items 1 to 5 of the MA-II:

$$
\text{Utility (MA-II-based)} = 0.4293 + (0.0336 \times \text{MA1}) + (0.0071 \times \text{MA2}) + (0.0053 \times \text{MA3}) + (0.0107 \times \text{MA4}) + (0.0001 \times \text{MA5})
$$

This reduced item-based model gave an acceptable goodness-of-fit in the development sample.
but showed better fit in the validation sample \((R^2 = 0.550)\). Thus, the model explains 55% of the EQ-5D variation. Mean and median values for MAPE were 20% and 7% (IQR 3% to 14%) in the development sample and 19% and 7% (IQR 3% to 14%) in the validation sample.

Mapping the SF-6D

A good correlation between the SF-6D and the MA-II scale was found (Pearson’s \(r = 0.734\), Spearman’s \(r = 0.741\)). When comparing the scale-based and the item-based regression model, goodness-of-fit as measured by \(R^2\) was comparably high for both models (0.573 vs. 0.547, Table 3). Again, there was no effect of center or language on the results, and plotting the regression residuals revealed no evidence of nonlinearity. In the model, which used the six single items of the MA-II, the items 4 and 6 were uninformative for the prediction of utility. Nevertheless, because item 4 was informative in the EQ-5D-based model,

\begin{table}
\centering
\begin{tabular}{lccc}
\hline
& Preoperative patients & Postoperative patients & Total sample \\
\hline
Recruitment site & & & \\
Frankfurt, Germany & 113 & 287 & 400 \\
Naples, Italy & 65 & 84 & 149 \\
Prague, Czechia & 106 & 94 & 200 \\
Barcelona, Spain & 0 & 53 & 53 \\
Málaga, Spain & 18 & 73 & 91 \\
Age (years) & 40.4 (10.8) & 41.5 (10.8) & 41.2 (10.8) \\
Sex (n females) & 224 (74.2%) & 478 (80.9%) & 702 (78.6%) \\
Body mass index (kg/m²) & 46.0 (8.4) & 34.9 (8.4) & 38.9 (9.9) \\
Diabetes mellitus & & & \\
Requiring oral drugs & 40 (13.2%) & 28 (4.7%) & 68 (7.6%) \\
Requiring insulin & 18 (6.0%) & 10 (1.7%) & 28 (3.1%) \\
Pulmonary disease & & & \\
Requiring drug therapy & 54 (17.9%) & 26 (4.4%) & 80 (9.0%) \\
Requiring CPAP treatment & 17 (5.6%) & 16 (2.7%) & 33 (3.7%) \\
Metabolic disease & & & \\
Laboratory parameters elevated & 75 (24.8%) & 38 (6.4%) & 113 (12.7%) \\
Requiring regular drug therapy & 44 (14.6%) & 18 (3.0%) & 62 (6.9%) \\
Cardiovascular disease & & & \\
Abnormal findings or blood pressure & 42 (14.0%) & 48 (8.1%) & 90 (10.1%) \\
Requiring regular drug therapy & 101 (33.4%) & 100 (16.9%) & 201 (22.5%) \\
Knee arthritis & & & \\
Requiring daily analgesics or surgery & 71 (23.5%) & 82 (13.9%) & 153 (17.1%) \\
Values are shown as means (SD) or counts (percentages). All differences between pre- and postoperative patients (except age) are statistically significant. CPAP, continuous positive airway pressure.
\hline
Table 2 Comparison of health-related quality-of-life between pre- and postoperative patients
\end{tabular}
\end{table}

\begin{table}
\centering
\begin{tabular}{lccc}
\hline
& Preoperative patients & Postoperative patients & Total sample \\
\hline
SF-36 health domain scales & & & \\
Physical functioning & 51.5 (28.8) & 75.3 (26.2) & 67.3 (29.3) \\
Role-physical & 50.8 (33.4) & 73.4 (30.0) & 65.8 (33.0) \\
Bodily pain & 49.8 (31.3) & 63.5 (32.1) & 58.9 (32.5) \\
General health & 52.4 (26.2) & 68.1 (24.1) & 62.8 (25.9) \\
Vitality & 53.0 (15.7) & 55.7 (16.5) & 54.8 (16.3) \\
Social functioning & 51.3 (24.2) & 60.9 (28.5) & 57.6 (27.5) \\
Role-emotional & 45.5 (35.7) & 60.5 (38.9) & 55.5 (38.5) \\
Mental health & 53.9 (32.8) & 57.3 (32.9) & 56.2 (32.9) \\
SF-6D utility score & 0.624 (0.117) & 0.733 (0.128) & 0.699 (0.134) \\
MA-II single items (1 to 10 scale) & & & \\
General mood (MA1) & 6.1 (2.5) & 7.7 (2.0) & 7.1 (2.3) \\
Physical function (MA2) & 4.9 (2.5) & 7.1 (2.4) & 6.4 (2.6) \\
Social relationships (MA3) & 6.3 (2.9) & 8.1 (2.1) & 7.5 (2.5) \\
Ability to work (MA4) & 6.3 (2.8) & 7.8 (2.3) & 7.3 (2.6) \\
Sexuality (MA5) & 5.6 (3.2) & 7.0 (2.7) & 6.5 (3.0) \\
Eating behavior (MA6) & 5.9 (2.3) & 7.5 (2.1) & 7.0 (2.3) \\
MA-II summary score & 35.4 (12.1) & 45.3 (9.8) & 41.9 (11.6) \\
EQ-5D utility score & 0.754 (0.217) & 0.869 (0.144) & 0.830 (0.180) \\
EQ-5D VAS utility & 58.3 (23.7) & 74.8 (19.3) & 69.3 (22.3) \\
\hline
\end{tabular}
\end{table}

Values are shown as means (SD). MA-II, Moorehead-Ardelt II questionnaire (6 to 60 scale). To convert to classical scale (–3 to 3) substract 6, divide by 9 and again substract 3. EQ-5D, EuroQol-5D; SF-6D, Short Form-6D; VAS, visual analog scale.
the preferred model includes the first five items of the MA-II:

Utility (MA-II-based) = 0.3325 + (0.0232 × MA1) + (0.0089 × MA2) + (0.0068 × MA3) + (0.0044 × MA4) + (0.0086 × MA5).

Applying this model to the development and validation sample, it explained 57.2% and 54.3% of the observed variation in SF-6D utilities (Table 4). Mean and median values for MAPE were 11% and 9% (IQR 4% to 14%) in the development sample and 11% and 8% (IQR 4% to 14%) in the validation sample.

Comparison of Predictive Models
A statistical comparison of the errors between observed and expected values (using MAPE data) confirmed the superiority of the EQ-5D-based mapping algorithm (P = 0.045 by Wilcoxon test, n = 732).

Table 3 Model selection for prediction of EuroQoL-5D (EQ-5D)-based and Short Form-6D (SF-6D)-based utility values (based on development sample)

<table>
<thead>
<tr>
<th>Specification of model</th>
<th>N patients</th>
<th>R²</th>
<th>α</th>
<th>Sₑ</th>
<th>βᵢ</th>
<th>SEᵢ</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D = α + β₁,MA-II + error</td>
<td>Scale-based</td>
<td>412</td>
<td>0.387</td>
<td>0.434</td>
<td>0.026</td>
<td>0.010*</td>
</tr>
<tr>
<td>EQ-5D = α + β₁,MA-II-1 + item-based</td>
<td>412</td>
<td>0.441</td>
<td>0.422</td>
<td>0.026</td>
<td>β₁ = 0.033*</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>+ β₂,MA-II-2</td>
<td></td>
<td></td>
<td></td>
<td>β₂ = 0.007</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>+ β₃,MA-II-3</td>
<td></td>
<td></td>
<td></td>
<td>β₃ = 0.005</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>+ β₄,MA-II-4</td>
<td></td>
<td></td>
<td></td>
<td>β₄ &lt; 0.001</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ β₅,MA-II-5</td>
<td></td>
<td></td>
<td></td>
<td>β₅ = 0.003</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>+ β₆,MA-II-6 + error</td>
<td></td>
<td></td>
<td></td>
<td>β₆ = 0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>EQ-5D = α + β₁,MA-II</td>
<td>Covariate-adjusted</td>
<td>412</td>
<td>0.394</td>
<td>0.529</td>
<td>0.066</td>
<td>β₁ = 0.009*</td>
</tr>
<tr>
<td></td>
<td>+ β₂,Sex</td>
<td></td>
<td></td>
<td></td>
<td>β₂ = -0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>+ β₃,MA-II + error</td>
<td></td>
<td></td>
<td></td>
<td>β₃ = 0.031</td>
<td>0.017</td>
</tr>
<tr>
<td>SF-6D = α + β₁,MA-II + error</td>
<td>Scale-based</td>
<td>367</td>
<td>0.547</td>
<td>0.321</td>
<td>0.019</td>
<td>0.009*</td>
</tr>
<tr>
<td>SF-6D = α + β₁,MA-II-1 + item-based</td>
<td>367</td>
<td>0.573</td>
<td>0.328</td>
<td>0.019</td>
<td>β₁ = 0.023*</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ β₂,MA-II-2</td>
<td></td>
<td></td>
<td></td>
<td>β₂ = 0.009*</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ β₃,MA-II-3</td>
<td></td>
<td></td>
<td></td>
<td>β₃ = 0.007*</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ β₄,MA-II-4</td>
<td></td>
<td></td>
<td></td>
<td>β₄ = 0.004</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ β₅,MA-II-5</td>
<td></td>
<td></td>
<td></td>
<td>β₅ = 0.008*</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>+ β₆,MA-II-6 + error</td>
<td></td>
<td></td>
<td></td>
<td>β₆ = 0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>SF-6D = α + β₁,MA-II</td>
<td>Covariate-adjusted</td>
<td>367</td>
<td>0.556</td>
<td>0.426</td>
<td>0.046</td>
<td>0.009*</td>
</tr>
<tr>
<td></td>
<td>+ β₂,Sex</td>
<td></td>
<td></td>
<td></td>
<td>β₂ = -0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>+ β₃,BMI + error</td>
<td></td>
<td></td>
<td></td>
<td>β₃ = -0.023*</td>
<td>0.011</td>
</tr>
</tbody>
</table>

*Statistically significant coefficient.
MA-II, Moorehead-Ardelt II questionnaire.
Discussion

Bariatric surgery is an emerging field in one of medicine’s most dynamic eras. Because widespread acceptance of this treatment option has come only recently, research on economic or HRQoL issues in morbidly obese patients is still in its infancy. The results of this study allow researchers to calculate preference-based utility values from an HRQoL instrument that is commonly used in bariatric surgery. By applying our algorithm to MA-II data from randomized trials, the utilities of two different treatment strategies could be assessed and compared.

The most critical aspect of a mapping algorithm is the goodness-of-fit it provides. In this study both algorithms achieved values of $R^2$ greater than 50% in the validation sample. This is comparable to other recent studies carried out on patients with obesity [23], prostate cancer [24], Crohn’s disease [25], and growth hormone deficiency [26]. In these studies $R^2$ values ranged between 0.29 and 0.69. Similarly, the MAPE values of the present study (7% to 9%) were in line with previous research on abdominal diseases [25]. Therefore, the relationship between MA-II and both SF-6D and EQ-5D appears to be strong enough for predictive purposes.

We decided against using the EQ-5D VAS as a target variable for modelling utility, because inconsistencies between VAS values and EQ-5D-based utilities are known to be present in about two-thirds of cases [27]. In the present study, VAS values were clearly lower and more spread out than EQ-5D utilities. Although the VAS allows for a direct valuation of health states, our results probably suggest that VAS data tend to be less valid and reliable, at least if respondents are assessing their own health status rather than valuing hypothetical article cases.

We consider the heterogeneity of the present cohort as an advantage of the study, because it increases the generalizability of the results. If recruitment was restricted to only one type of bariatric procedure, this would possibly prevent the usage of the algorithm in other patient cohorts. For example, European surgeons traditionally have preferred gastric banding over gastric bypass, although American surgeons do the opposite [28]. In this regard, the cohort we studied is representative of the general population of morbidly obese patients, including those who received surgery.

Although previous research has used utility measures primarily on overweight rather than obese patients [23,29,30], the BMI for patients in this study ranged to extreme values. The fact that BMI was uninformative for the regression model supports the assertion that all estimated parameters are robust over the full range from normal weight to morbid obesity. Furthermore, the differences between the development and the validation sample were relatively small. The majority of the estimated coefficients remained unchanged when additional parameters were added to the model, which underpins the validity of the statistical model. The sole reliance on a essentially linear regression model could still be considered a shortcoming of the present study. In other mapping studies, however, the use of additional methods (Jackknife and bootstrap analysis) yielded very similar results as compared to multiple regression analysis.

Although the MA-II is well accepted by the surgical community, this study also detected some inherent problems of this questionnaire. In the prediction of utility values not all of the items were found to be significantly associated with EQ-5D and SF-6D-based score values. Not unexpectedly, the more generic items of the MA-II were highly correlated to generic HRQoL and utility values. On the other hand, the most disease-specific sixth item of the MA-II had to be removed from the estimation algorithm. This can be possibly explained by the fact that many patients see their change in eating habits not as an improvement but as a restriction in HRQoL, because they were (to some extent) enjoying their supranormal preoperative food intake. Another possible explanation is that a healthy attitude to food truly is an aspect of HRQoL, but both
EQ-5D and SF-6D fail to record it. Further validation studies on the MA-II need to be performed to examine how well this item is able to discriminate between patients with high and low HRQoL. It also is necessary to assess the predictive properties of other measures that can be used in obese patients.

It is known that the method of data collection has an impact on the responses. Usually, HRQoL data tend to be more favorable when data are collected by personal interviews. On the other hand, patients wishing bariatric surgery might tend to aggravate their symptoms in psychometric tests to avoid being denied surgery. In this study, patients completed the questionnaires themselves in the outpatient clinic. Patients were explicitly informed that their responses would not affect their health care in any manner. This reduces the potential for biased responses. Even if the way of data collection caused some bias, the bias would affect all questionnaires in more or less the same way and therefore it would not affect the correlation between the questionnaires.

The question which of the two utility instruments should be used in the estimations is not easy to answer, although median estimation error was slightly smaller for the EQ-5D-based than for the SF-6D-based regression models. From previous comparison studies it is known that the SF-6D can better describe nearly normal health states [31,32], such as in patients with mild residual overweight and no further health problems. Our results are supportive of mild ceiling effects of the EQ-5D (Fig. 1). Because the EQ-5D was also better accepted by patients, we would recommend applying the EQ-5D-based mapping algorithm. Alternatively, the use of an MA-II scale-based algorithm also would provide reliable estimates of utility values.

Principally, it appears preferable to use a preference-based measure such as the EQ-5D or the SF-6D rather than a disease-specific instrument. Ideally, the present study would also have included further direct utility measures such as standard gamble or time trade-off [27]. Nevertheless, the sensitivity of generic instruments may be lower [17,33]. Second, a recent review of derivation studies found that disease-specific instruments allow calculation of utilities with similar precision as generic instruments [34]. Third, the application of nondisease-specific instruments is not possible in many clinical settings. In these situations, it is valid to estimate utilities directly from MA-II items, since this study has detected only a small risk of error for this algorithm. The EQ-5D-based algorithm seems to be preferable, mainly because of the generally better psychometric characteristics of this measure in patients with mild or moderate disease, but this needs to be confirmed in future studies. According to the distribution of the sample, we believe that the utility model is suitable for all patients with a BMI greater than 30 and those who have undergone bariatric surgery.

The authors thank Dr. Melodie K. Moorehead, Prof. Martin Fried, Dr. Amador García Ruiz de Gordejuela, Dr. Pier Paolo Cutolo, Dr. Melanie Afiüpper-Fink, Dr. Brigitte Habermalz, Dr. Antoine Mpabansi, Dr. Saeed Khoursandi, and Eva Häusler for their help in designing and executing this study.

Source of financial support: This study was funded by Ethicon Endo-Surgery (Europe), GmbH. The first author retained full control over methodology and had unrestricted rights to publish the results.

References


23 Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. Value Health 2004;7:490–8.


