Stolz, Maike et al.

Article — Published Version

Assessment of health-related quality of life in individuals with depressive symptoms: validity and responsiveness of the EQ-5D-3L and the SF-6D

The European Journal of Health Economics

Provided in Cooperation with:
Springer Nature

Suggested Citation: Stolz, Maike et al. (2022) : Assessment of health-related quality of life in individuals with depressive symptoms: validity and responsiveness of the EQ-5D-3L and the SF-6D, The European Journal of Health Economics, ISSN 1618-7601, Springer, Berlin, Heidelberg, Vol. 24, Iss. 8, pp. 1297-1307, https://doi.org/10.1007/s10198-022-01543-w

This Version is available at:
https://hdl.handle.net/10419/309892

The European Journal of Health Economics (2023) 24:1297–1307
https://doi.org/10.1007/s10198-022-01543-w

ORIGINAL PAPER

Assessment of health-related quality of life in individuals with depressive symptoms: validity and responsiveness of the EQ-5D-3L and the SF-6D

Maike Stolz[1,2] · Christian Albus[3] · Manfred E. Beutel[4] · Hans-Christian Deter[5] · Kurt Fritzsche[6] · Christoph Herrmann-Lingen[7,8] · Matthias Michal[4] · Katja Petrowski[9] · Joram Ronel[10,11] · Jobst-Hendrik Schultz[12] · Wolfgang Söllner[13] · Cora Weber[14] · Martina de Zwaan[15] · Christian Krauth[1,2]

Received: 24 May 2022 / Accepted: 24 October 2022 / Published online: 16 November 2022

Abstract

Background The EQ-5D and the SF-6D are examples of commonly used generic preference-based instruments for assessing health-related quality of life (HRQoL). However, their suitability for mental disorders has been repeatedly questioned.
Objective To assess the responsiveness and convergent validity of the EQ-5D-3L and SF-6D in patients with depressive symptoms.
Methods The data analyzed were from cardiac patients with depressive symptoms and were collected as part of the SPIRR-CAD (Stepwise Psychotherapy Intervention for Reducing Risk in Coronary Artery Disease) trial. The EQ-5D-3L and SF-6D were compared with the HADS (Hospital Anxiety and Depression Scale) and PHQ-9 (Patient Health Questionnaire) as disease-specific instruments. Convergent validity was assessed using Spearman's rank correlation. Effect sizes were calculated and ROC analysis was performed to determine responsiveness.
Results Data from 566 patients were analysed. The SF-6D correlated considerably better with the disease-specific instruments ($|r_s|$= 0.63–0.68) than the EQ-5D-3L ($|r_s|$= 0.51–0.56). The internal responsiveness of the SF-6D was in the upper range of a small effect (ES: − 0.44 and − 0.47), while no effect could be determined for the EQ-5D-3L. Neither the SF-6D nor the EQ-5D-3L showed acceptable external responsiveness for classifying patients' depressive symptoms as improved or not improved. The ability to detect patients whose condition has deteriorated was only acceptable for the EQ-5D-3L.
Conclusion Overall, both the convergent validity and responsiveness of the SF-6D are better than those of the EQ-5D-3L in patients with depressive symptoms. The SF-6D appears, therefore, more recommendable for use in studies to evaluate interventions for this population.

Keywords EQ-5D-3L · SF-6D · Mental disorder · Validity · Responsiveness

JEL Classification I100

Background

Depressive disorders are of great importance to society due to their burden of disease, prevalence, frequent recurrence or long-lasting course, increased use of the health care system and the associated direct and indirect costs [1]. In terms of disability adjusted life years (DALYs), the burden of disease of depressive disorders was in third place in a worldwide comparison of all illnesses in 2001 in high-income countries [2]. According to a prognosis by the World Health Organisation (WHO), depressive disorders will be the most significant of the widespread diseases that impair or shorten life by the year 2030. Since the years of life lost due to premature death are of little significance in depression, it becomes clear how severely the way of life is impaired by this illness [2, 3]. For the individual concerned, the presence of depressive symptoms is associated with a loss of health-related quality of life (HRQoL) by influencing the physical, emotional and social aspects of well-being [4, 5].

✉ Maike Stolz
stolz.maike@mh-hannover.de

Extended author information available on the last page of the article

The EQ-5D and SF-6D are generic multi-attribute health status classification systems, which are used to assess HRQoL in health economic evaluations [6, 7]. By evaluating health states according to their relative value (derived from preferences), and summarizing them into a single index value (utility value). They are a widely used because of being an indirect alternative for measuring preferences using simple questionnaires, as measuring preferences through direct questioning and assessment by the patient concerned is very time-consuming and complex. [8–11].

Results of health economic evaluations are part of allocation decisions of limited resources in the health care system. A prerequisite for a reliable comparison of different interventions and a resulting "fair" allocation is the suitability of the health economic quality-of-life instruments in the context of different diseases and populations [8].

However, the suitability of generic instruments for assessing HRQoL in mental disorders has repeatedly been questioned [12–18]. The main concerns are based on the design of these instruments with a focus on physical complaints, so that (changes in the) psychological components are not sufficiently taken into account in the summary scores and the index scores [15–18]. This seems to be especially true for the EQ-5D, as four of the five dimensions are in the physical domain, while the six dimensions of the SF-6D are balanced between the physical and psychological domains [19]. In general, it is often discussed that responsiveness of generic instruments is lower than that of disease-specific instruments because the questions are less specific to the symptoms of the underlying disease and therefore minor changes are not captured. However, the generality of this statement is controversial [8, 20, 21].

The purpose of this study was to evaluate whether the EQ-5D-3L and SF-6D, as examples of commonly used generic preference-based instruments for assessing HRQoL, are suitable for patients with depressive symptoms and whether either instrument is superior to the other for this purpose. To assess the responsiveness and convergent validity of the EQ-5D-3L and SF-6D, they were compared to the depression scales of the disease-specific Hospital Anxiety and Depression Scale (HADS) and Patient Health Questionnaire (PHQ-9). The following hypotheses were examined:

(1) The correlation between the EQ-5D-3L and disease-specific instruments differs from the correlation between the SF-6D and disease-specific instruments.
(2) The responsiveness of the generic instruments differs from the responsiveness of the disease-specific instruments.
(3) There is a difference in the responsiveness of the EQ-5D-3L and SF-6D.

Methods

The analyses carried out are based on data from the Stepwise Psychotherapy Intervention for Reducing Risk in Coronary Artery Disease (SPIRR-CAD) study. Details and results of the randomized controlled trial are described elsewhere [22, 23]. Briefly, the SPIRR-CAD study was designed to test the hypothesis that a stepwise psychotherapy intervention is more effective in mitigating depressive symptoms in cardiac patients than one information session added to usual care. Inclusion criteria included age between 18 and 75 years, documented coronary artery disease (CAD) and a depression score higher than 7 points on the HADS depression scale. Exclusion criteria included inability to speak German, severe heart failure (New York Heart Association (NYHA) Class IV), scheduled cardiac surgery within the next 3 months, severe depressive episode according to the Structured Clinical Interview for DSM-IV or other severe or life-threatening physical or mental illness. All patients received usual care by their general practitioner and/or cardiologist. Patients in the control group additionally received one information session of 30 to 45 min providing information about healthy behaviours and psychosocial factors in CAD. Patients in the intervention group were offered three individual psychotherapy sessions. All patients were reassessed with the HADS depression scale, and only those continue to show depressive symptoms (HADS score > 7) after 4 to 8 weeks were offered 25, 90-min sessions of group psychotherapy.

Instruments

Various survey instruments were used in the SPIRR-CAD study. The SF-6D (SF-36), EQ-5D-3L, HADS and PHQ-9 were available for the comparison of generic and disease-specific instruments in depressive disorders.

The HADS depression scale is a psychometric self-assessment tool to measure depressive symptoms in patients with primary somatic diseases [24, 25]. It consists of seven items each rated from 0 to 3 according to severity of difficulty experienced. Total score ranges from 0 (no depression) to 21, in which ≤ 7 points are considered unremarkable, 8–10 points are considered reflecting marginal depression and ≥ 11 points are considered conspicuous.

The Patient Health Questionnaire 9-item (PHQ-9) is a self-assessment depression screening tool for administration among adults in primary care settings [26, 27]. It consists of nine items each rated from 0 to 3 according to frequency of occurrence. Total score ranges from 0 (no depression) to 27, in which ≤ 4 points are assessed as no depressive symptoms, 5–9 points as mild or moderate depressive symptoms and ≥ 10 points as suggestive of major depression.

The SF-6D is a generic preference-based index instrument, developed for use in health economic evaluation studies. It can be derived from data from the SF-36, which is one of the most widely used generic HRQoL instruments worldwide [6]. The SF-6D consists of eleven items (of the SF-36) that are divided into six dimensions: physical functioning, role limitations, social functioning, pain, mental health and vitality. Each dimension has between two and six levels. A SF-6D health state is defined by selecting one level from each dimension resulting in 18,000 different possible health states. In the end, every health state can be described by an index value. Therefore, a representative sample of the general population has to assess selected health states using preference-based methods (e.g. standard gamble or time trade off). A value set, weighting the levels in each dimension, is calculated from the results using multiple regression analyses. This value set can be used to calculate a single index value out of the data derived from an applied SF-36 questionnaire.

The EQ-5D is a generic preference-based index instrument for describing, quantifying and valuing HRQoL [7]. It comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension of the EQ-5D-3L used in this study has three levels. An EQ-5D-3L health state is defined by selecting one level from each dimension resulting in 243 different possible health states [28]. A single index value can then be derived from the characterized health state using the method described above for the SF-6D.

The choice of a value set can affect the resulting index value, because value sets are generally meant to reflect the preferences of specific countries which can be different from each other [29]. As there are no values for the German context, it was decided to use the British tariff for the SF-6D ($SF\text{-}6D_{UK}$) [6]. Although there is a German value set for the EQ-5D-3L ($EQ\text{-}5D\text{-}3L_{GER}$) [30], the main analyses to test hypotheses were carried out using the UK tariff ($EQ\text{-}5D\text{-}3L_{UK}$) [31, 32] for two reasons. (1) A comparison of the suitability of the EQ-5D-3L and SF-6D for people with depressive symptoms is more valid when the underlying preferences are from the population of the same country. (2) The German tariff does not contain a discount value for mild depression. This means that when comparing responsiveness, a change between mild and non-existent depression cannot be mapped. However, since the German tariff for the EQ-5D-3L should not be completely ignored in the context of a German study population, the results of the analyses are also presented for the $EQ\text{-}5D\text{-}3L_{GER}$.

Data analysis

All analyses were carried out using SPSS version 22 and cocor web interface [33].

To investigate whether the EQ-5D-3L and SF-6D are appropriate for use with patients with depressive symptoms, convergent validity and responsiveness were examined.

The convergent validity describes the degree to which two measures of constructs that theoretically should be correlated, are in fact correlated [34]. In this analysis the index scores of the EQ-5D-3L indices and $SF\text{-}6D_{UK}$ were compared to the sum scores of the disease-specific HADS and PHQ-9. Spearman rank correlation coefficients (r_s) of the scores at 6 months were calculated to build a correlation matrix.

To test hypothesis 1, correlation coefficients of the $EQ\text{-}5D\text{-}3L_{UK}$, $SF\text{-}6D_{UK}$, HADS and PHQ-9 have to be compared and a test of significance was necessary to control for possible differences occurring by chance [33, 35]. Dunn and Clark's z was chosen because of its appropriateness for dependent correlations with either overlapping or nonoverlapping variables [33, 36, 37]. The analyses were carried out using the cocor web interface [33].

Responsiveness is defined as the ability of an instrument to detect change over time [38]. Internal responsiveness was assessed using effect sizes [39]. Since there is controversy regarding the most appropriate effect size for calculating responsiveness [39], both the standardized effect size (SES) and the standardized response mean (SRM) were used:

$SES = \frac{M_{t1}-M_{t2}}{SD_{t1}}$ and $SRM = \frac{M_{t1}-M_{t2}}{SD_{t2-t1}}$,where "$M_{t1}$" is the arithmetic mean at baseline assessment, "M_{t2}"" is the arithmetic mean at 18 months, "SD_{t1}" is the standard deviation (SD) at baseline assessment and SD_{t2-t1} is the standard deviation (SD) of the measured difference between baseline assessment and assessment after 18 months. A clinically relevant change of at least two points on the HADS is an indicator of an improvement/deterioration and is used as a reference criterion [40].

So far, there are no specific benchmarks for effect sizes as a measure of responsiveness. For this reason, the "rule of thumb" according to Cohen is often used to assess effect sizes in intervention studies [41]. This means that a value between 0.2 and 0.49 corresponds to a small effect, a value between 0.5 and 0.79 corresponds to a medium effect and a value of > 0.8 corresponds to a large effect [42–44].

To test hypotheses 2 and 3 the Modified Jacknife Test was used as a test of significance [45–47]. This test is based on a linear regression, where the dependent variable contains the difference of the SES/SRM between the two instruments to be compared, while the independent variable consists of the "centered SES/SRM". The "centered SES/SRM" is formed by subtracting the mean SES/SRM of one of the two instruments to be compared (which one is not relevant) from the individual SES/SRM for each patient. A significant intercept coefficient represents a significant difference between the SES/SRM of the two scales to be compared [46, 47].

To control for possible violation of requirements for linear regression (normal distribution of the residues and homoscedasticity) the bootstrap method was carried out. The intercept coefficient and the associated Modified Jacknife *Test* are only considered significant if the confidence intervals generated by bootstrapping do not contain the value "0".

External responsiveness was assessed using Receiver operating characteristics (ROC) curves and the area under the curve (AUC) as a reference number [39, 48, 49]. In this context, responsiveness is described in terms of change sensitivity and change specificity. "Change sensitivity" means the probability of the instrument correctly classifying patients who demonstrate a change on an external criterion, whereas "change specificity" means the probability of an instrument correctly classifying patients who do not demonstrate change on an external criterion. A change can either mean an improvement or a deterioration. Separate ROC curves must be calculated for both cases [49]. As an external criterion for change, the HADS was used. Here, a change by two points between baseline assessment and follow-up assessment after 18 months was defined as the Minimal Clinical Important Difference (MCID) [40] and accordingly assigned to the status "changed". If there was a difference of less than two points, the status was considered as being "unchanged".

The AUC represents the probability that an instrument correctly classifies patients as improved or not improved and deteriorated or not deteriorated, respectively [39, 48]. An AUC of 0.5 means that an instrument cannot discriminate between patients whose status has changed and Patients whose status has not changed, while a value of 1.0 corresponds to perfect discriminatory power. A value ≥ 0.7 is considered moderate [50].

The formulated hypotheses are global hypotheses, which have to be proved by multiple statistical tests. To avoid the error of multiple comparisons a Bonferroni correction was conducted. Each single test was evaluated with a corrected α level (α') with

$$\alpha' = \alpha / k,$$

where 'α' is the critical probability (p) level and 'k' is the number of tests performed [51]. The assumed α level of 0.05 for each single test was therefore corrected to an α' of 0.0125 for hypothesis 1 (for $k=4$), to an α' of 0.00625 (for $k=8$) for hypothesis 2 and to an α' of 0.025 (for $k=2$) for hypothesis 3 (see Supplementary Information).

Results

The cohort consisted of 566 patients whose detailed characteristics have been reported elsewhere [23]. The mean age was 59.2 years and 21.1% were female. Most patients (81.7%) were classified in NYHA class I or II. Overall, 11.6% of the patients received antidepressant medication and 11.1% were in psychotherapy within the preceding 12 months.

Descriptive analysis

Means and medians of the compared instruments are shown in Table 1. With a mean score of 10.42 on the baseline measure on the HADS and 9.95 on the PHQ-9, participants in the study had mild to moderate depressive symptoms on average. The mean index value of the SF-6D$_{UK}$ was at least 0.03 points lower than that of the EQ-5D-3L$_{UK}$ at all three measurement points. In addition, the standard deviation of the SF-6D$_{UK}$ was only half as large as that of the EQ-5D-3L$_{UK}$. Noticeably, the mean value of the EQ-5D-3L$_{UK}$ was significantly lower than that of the EQ-5D-3L$_{GER}$, with a difference of at least 0.13 points.

Medians and means were close for most instruments. Clear differences can only be seen in the EQ-5D-3L$_{UK}$ and EQ-5D-3L$_{GER}$. The difference between the SF-6D$_{UK}$ and the EQ-5D-3L$_{UK}$ was more obvious when looking at the median, with differences between 0.08 and 0.11 points, than for the mean differences.

Convergent validity

A higher score on the generic instruments equates to a better state of health, whereas a higher score on the disease-specific instruments is associated with a more severe disorder. As expected, this results in a positive correlation between the generic and disease-specific instruments among each other and a negative correlation between the generic and disease-specific instruments (Table 2). The SF-6D$_{UK}$ correlates best with the EQ-5D-3L$_{UK}$, while the EQ-5D-3L$_{UK}$ is more strongly associated with the EQ-5D-3L$_{GER}$. Overall, the SF-6D$_{UK}$ correlates considerably better with the disease-specific instruments ($|r_s|$= 0,63–0,68) than the EQ-5D-3L$_{UK}$ ($|r_s|$= 0,51–0,56) or EQ-5D-3L$_{GER}$ ($|r_s|$= 0,42–0,45). The comparison of the correlation coefficients of the SF-6D$_{UK}$ with the disease-specific instruments and the EQ-5D-3L$_{UK}$ with the disease-specific instruments could confirm that the differences found were significantly different and in favour of the SF-6D (see Supplementary Information, Table 1). Therefore, it can be assumed that the SF-6D$_{UK}$ shows a higher convergent validity for use in people with depressive symptoms than the EQ-5D-3L$_{UK}$, which confirms hypothesis 1.

Internal responsiveness

The SF-6D$_{UK}$ reached values between − 0.44 and − 0.47 in the upper range of a small effect, while no effect can be

Table 1 Means and medians of individual instruments at three central measurement points

		Baseline	6 months	18 months
EQ-5D-3L$_{GER}$	*n*	521	436	384
	Mean (SD)	0.79 (0.22)	0.80 (0.23)	0.81 (0.24)
	Median (IQR)	0.89 (0.79-0.89)	0.89 (0.79-0.89)	0.89 (0.79-1.00)
EQ-5D-3L$_{UK}$	*n*	521	436	384
	Mean (SD)	0.64 (0.26)	0.67 (0.28)	0.68 (0.28)
	Median (IQR)	0.71 (0.62-0.80)	0.73 (0.62-0.80)	0.73 (0.66-0.85)
SF-6D$_{UK}$	*n*	475	425	355
	Mean (SD)	0.60 (0.11)	0.64 (0.12)	0.65 (0.12)
	Median (IQR)	0.60 (0.53-0.65)	0.62 (0.56-0.73)	0.64 (0.56-0.74)
HADS	*n*	566	445	397
	Mean (SD)	10.42 (2.54)	8.98 (3.90)	8.13 (3.94)
	Median (IQR)	10.00 (8.00-12.00)	9.00 (6.00-12.00)	8.00 (6.00-11.00)
PHQ-9	*n*	526	446	387
	Mean (SD)	9.95 (5.27)	9.13 (5.03)	8.29 (5.15)
	Median (IQR)	9.00 (6.00-13.00)	9.00 (5.00-13.00)	7.00 (4.00-11.00)

n sample size, *SD* standard deviation, *IQR* interquartile range

Table 2 Correlation matrix of all instruments at T2

		EQ-5D-3L$_{GER}$	EQ-5D-3L$_{UK}$	SF-6D$_{UK}$	HADS	PHQ-9
EQ-5D-3L$_{GER}$	n	436	436	405	425	428
	r_s	1.00	0.93*	0.65*	− 0.42*	− 0.45*
EQ-5D-3L$_{UK}$	n		436	405	425	428
	r_s		1.00	0.72*	− 0.51*	− 0.56*
SF-6D$_{UK}$	n			425	412	416
	r_s			1.00	− 0.63*	− 0.68*
HADS	n				445	436
	r_s				1.00	0.71*
PHQ-9	n					449
	r_s					1.00

n sample size, r_s Spearman rank correlation coefficient

*p < 0.01

demonstrated for the EQ-5D-3L indices, equivalent to a non-existent sensitivity to change for the studied population (Table 3). Using the Modified Jacknife Test, it was possible to determine that the differences in responsiveness were significant, confirming that the SF-6D$_{UK}$ is more sensitive to change in depressive symptoms than the EQ-5D-3L$_{UK}$ (see Supplementary Information, Table 2). The picture was also heterogeneous for the disease-specific instruments. While the HADS was the most responsive instrument with medium to large effects, only small effects could be achieved for the PHQ-9 (0.31–0.36), which were even smaller than those of the SF-6D. Since all multiple comparisons between the generic and disease-specific instruments were significant, it must be stated that neither the generic nor the disease-specific instruments can be classified as being generally more responsive than the others (see Supplementary Information, Table 2).

External responsiveness

According to our current data, the ability to discriminate between patients who improved and those who did not improve cannot be considered as good or moderate for any of the here applied instruments (Table 4). Only the PHQ-9 can be classified as acceptable for detecting patients whose condition has improved based on the result of the HADS. The ability to detect patients whose condition had deteriorated was only acceptable for the EQ-5D-3L$_{UK}$.

Table 3 SES and SRM of the instruments

	SES	SRM
EQ-5D-3L$_{GER}$	− 0.09	− 0.08
EQ-5D-3L$_{UK}$	− 0.16	− 0.15
SF-6D$_{UK}$	− 0.47	− 0.44
HADS	0.90	0.62
PHQ-9	0.31	0.36

Negative values represent an improvement for the generic instruments, whereas positive values represent an improvement for the HADS and PHQ-9

Table 4 Area under the curve of the inserted instruments

	Improvement AUC [95% CI]	Deterioration AUC [95% CI]
EQ-5D-3L$_{GER}$	0.533 [0.462; 0.605]	0.587 [0.493; 0.682]
EQ-5D-3L$_{UK}$	0.553 [0.482; 0.625]	0.626 [0.533; 0.718]
SF-6D$_{UK}$	0.592 [0.524; 0.660]	0.572 [0.462; 0.682]
PHQ-9	0.634 [0.569; 0.699]	0.531 [0.431; 0.630]

AUC area under the curve, *CI* confidence interval

The ROC curves of improvement are based on $n = 334$ ($n = 221$ improved, $n = 113$ unchanged)

The ROC curves of deterioration are based on $n = 176$ ($n = 63$ deteriorated, $n = 113$ unchanged)

Figure 1 shows an example of the ROC curves of the EQ-5D-3L$_{UK}$.

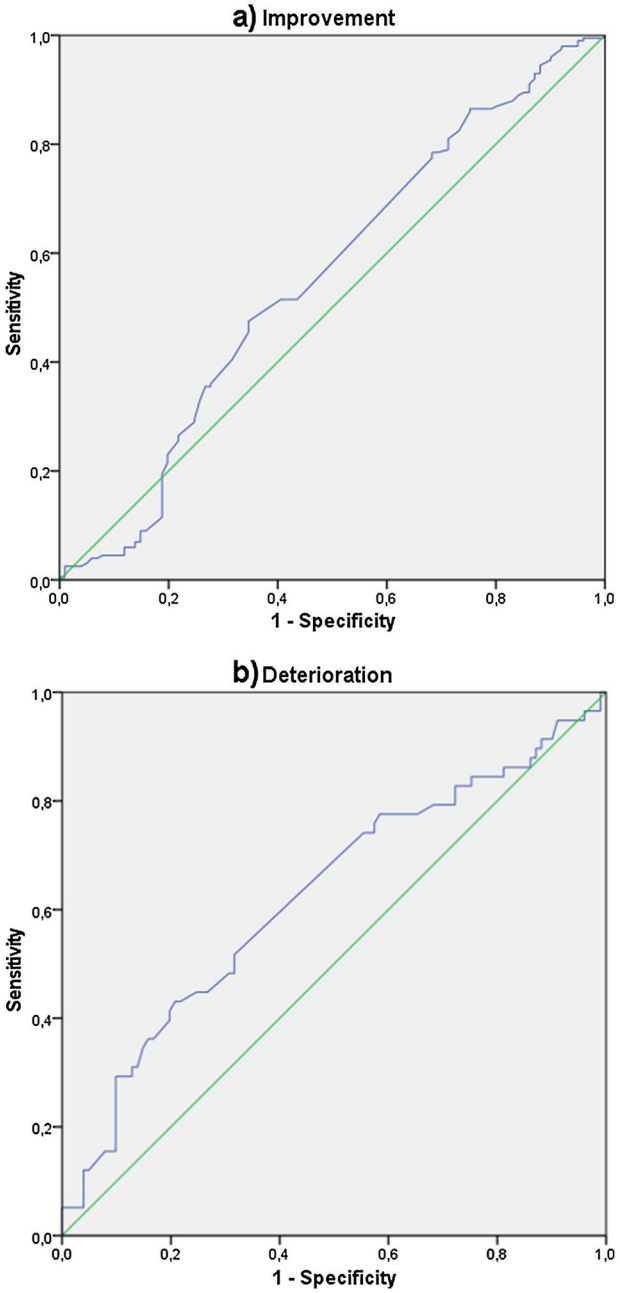

Fig. 1 ROC curves of the EQ-5D-3L$_{UK}$ for changes from T0 to T3

Discussion

In the light of limited resources in the health care system and constantly rising costs resulting from demographic change as well as expensive innovations in modern and advanced health care, there is a need for improving conditions for evidence-based allocation decisions. Health economic methods can help to make the allocation of resources in the healthcare system quantifiable. The use and suitability of generic preference-based quality of life instruments is essential for a targeted evaluation of medical interventions. However, its use in assessing mental disorders is questioned. In this context, the EQ-5D-3L and SF-6D were tested for use in patients with depressive symptoms.

Descriptive analysis

The mean baseline values for the SF-6D$_{UK}$ and for the EQ-5D-3L$_{UK}$ are comparable to those of other studies addressing depressive disorders that also used the British value sets used in this work. Sobocki et al. found an EQ-5D-3L$_{UK}$ score of 0.60 for mild depression in their observational study of medicated depressed patients [52]. When comparing the EQ-5D-3L$_{UK}$ and the SF-6D$_{UK}$ using data from a multicenter RCT to evaluate different therapeutic approaches for depressive and anxiety disorders, both the EQ-5D-3L$_{UK}$ and the SF-6D$_{UK}$ for mild symptoms were 0.60 [53]. A further comparison showed an index of 0.62 for the EQ-5D-3L$_{UK}$ and 0.63 for the SF-6D$_{UK}$ for people with mild depression in a population sample in Canada [54].

Similar to the present work, Lamers et al. found that the mean and median for the EQ-5D-3L$_{UK}$ differed significantly, while they were perfectly on top of each other for the SF-6D$_{UK}$ [53]. This discrepancy can possibly be explained by the preference values of the individual health conditions. For the SF-6D$_{UK}$, the worst health state is associated with a preference value of 0.296, while for the EQ-5D-3L$_{UK}$, negative values (for health states considered worse than death) and a preference value of -0.594 in the worst case are also possible [6, 32]. Such "outliers" lead to inaccurate estimates of mean values, which is also reflected in the significantly larger standard deviation of the EQ-5D-3L$_{UK}$ compared to the SF-6D$_{UK}$ [53]. A lower standard deviation enables more precise estimates. This is particularly relevant if quality-adjusted life years (QALYs) for cost-effectiveness studies are calculated based on the index values, which in turn can be used to compare two interventions and influence allocation decisions [55].

Convergent validity

The correlation between the HADS and the PHQ-9 for baseline measurement was in a similar range as in the study by Cameron and colleagues. The authors compared the two disease-specific instruments for use in primary care of patients with mild to moderate mental health problems and found a correlation coefficient of 0.68 [56].

A higher convergent validity was determined for the SF-6D$_{UK}$ and the depression scales than for the EQ-5D-3L$_{UK}$ as well as the EQ-5D-3L$_{GER}$ and the depression scales (Table 2). To our knowledge, this is the first direct comparison of the convergent validity of the SF-6D and EQ-5D-3L in a population of patients with depressive symptoms. The picture that emerges in the literature from studies comparing either the SF-6D or the EQ-5D-3L with disease-specific instruments is rather heterogeneous. Brazier et al. summarized the existing evidence for mild to moderate depression and found values between $|r_s|=0.35$ and $|r_s|=0.45$ for the relationship between the EQ-5D-3L and HADS and values between $|r_s|=0.56$ and $|r_s|=0.62$ for the relationship between the SF-6D and the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) [17]. Peasgood et al. found that the EQ-5D-3L correlated well with measures of depression severity ($|r_s|=0.54$–0.77) [57]. In a recent study, the convergent validity of the SF-6D was evaluated using the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) for depressive disorders. Good construct validity could be confirmed ($|r_s|=0.74$) [58].

Two possible explanations for the significantly better correlation of the SF-6D with disease-specific instruments compared to the correlation of the EQ-5D-3L with disease-specific instruments can be derived from the construction of the instruments. (1) The higher number of levels for the psychological dimension(s) of the SF-6D allows a more differentiated assessment of health states and particularly mild symptoms to be recorded more easily. However, since the low number of levels present in the EQ-5D-3L is a general problem that does not apply only to the psychological dimensions, the EuroQol group has since released an expanded version of the EQ-5D with five answer options (EQ-5D-5L). Abdin et al. investigated the convergent validity of the EQ-5D-5L using the Q-LES-Q for depressive disorders and determined good validity ($|r_s|=0.67$) [58]. Further research is needed to find out whether the validity of the EQ-5D-5L is really better than that of the EQ-5D-3L in depressed patients. (2) The EQ-5D focuses predominantly on the physical scope (four out of five dimensions), in contrast to the SF-6D, which is balanced between the physical and psychological scope with three dimensions each [19]. The weaker correlation of the EQ-5D-3L with the depression scales seems almost a logical consequence.

The predominant focus of the EQ-5D on the physical dimension of health might suggest that it is more suitable for use in somatic diseases than the SF-6D. Garcia-Gordillo and colleagues compared both instruments in a population of Parkinson's patients and found almost identical, strong correlations with a disease-specific questionnaire [59]. A similar picture emerged for rheumatic diseases, with a slight advantage for the SF-6D ($|r_s|=0.70$ vs. $|r_s|=0.80$) [60]. In contrast, the EQ-5D-3L was shown to be more suitable for multiple sclerosis and non-specific back pain [61, 62]. The convergent validity of the EQ-5D-3L, therefore appears to be equally good or in some cases even better than that of SF-6D for somatic diseases.

Internal responsiveness

The responsiveness of an instrument is of particular relevance in the context of health economic evaluations. If a HRQoL instrument is not responsive, a small but potentially clinically relevant change will not be reflected in the preference values and consequently not in the utility values (e.g. QALYs). Consequently, allocation decisions could be incorrectly influenced.

The SF-6D$_{UK}$ is significantly more responsive in a population with depressive disorders than the EQ-5D-3L$_{UK}$. The difference in responsiveness between the SF-6D$_{UK}$ and EQ-5D-3L$_{UK}$ is almost entirely due to the more than twice as high standard deviation of the EQ-5D-3L$_{UK.}$ A possible explanation from the different construction of the instruments has already been presented in the discussion on convergent validity. For both the EQ-5D-3L and the SF-6D, the responsiveness determined here was worse than that described in the literature for depressed or generally mentally ill people [63]. One possible reason for this could be the comorbid condition of the population, which influences

the results of the generic index instruments differently than the sole presence of a mental illness.

For a population of mildly to moderately depressed patients, the literature describes effect sizes between − 0.68 and − 1.05 for the HADS, which are comparable to the present sample [17]. It is thus significantly more sensitive to change than the PHQ-9, for which only a small effect could be demonstrated. When comparing the disease-specific and generic instruments, the HADS was always the more responsive instrument. In contrast, the SF-6D$_{UK}$ was significantly more responsive than the PHQ-9. The generally poorer responsiveness of the generic instruments, which has been repeatedly formulated but is also controversially discussed [8, 20, 21], could not be completely confirmed or refuted.

External responsiveness

None of the AUCs generated from the ROC analysis reached the threshold of 0.70, which would be equivalent to a moderate ability to discriminate between patients with changed and unchanged depressive symptoms. In the context of mental disorders, only one other study was identified that found similarly poor AUCs for the EQ-5D-3L and the SF-6D in patients with schizophrenia [64]. However, the result of this study must be viewed critically, since only one ROC curve was generated for general change and not separately for improvement and deterioration as recommended in the literature [48, 49].

An important factor influencing the AUC is the choice of the external indicator criterion. In the present study, a change by the MCID of two points on the HADS was chosen for this purpose, as this was used as the primary outcome in the SPIRR-CAD study. In addition, the HADS is used in particular for the assessment of psychological stress in the context of somatic illnesses and thus seems suitable for the present population of patients with CAD and depressive symptoms. This choice might have been problematic for the determination of the external responsiveness of generic instruments. Generic instruments map not only the psychological dimension and its changes, but also those of the other components of quality of life that are likely to be influenced by CAD and other somatic comorbidities of the population under study. All dimensions influence the index value, which leads to an expected greater variance than on the HADS. However, in the absence of a gold standard for recording the HRQoL of mentally ill people, this is a general problem in quality of life research [17, 63].

General aspects

Based on the results discussed so far, the SF-6D appears to be more valid and responsive than the EQ-5D-3L in patients with depressive symptoms. In addition to being more suitable for this specific population, the SF-6D as being derived from the SF-36 has another advantage. As a profile instrument, the SF-36 offers a detailed description of the individual dimensions of HRQoL and is therefore able to assess the consequences of an intervention in detail. The SF-36 is also widespread and often used in efficacy studies [8]. With its direct derivation from the SF-36 (or SF-12), the SF-6D offers the possibility to create a preference-based index value for cost–benefit analyses in the context of an effectiveness study without the need for an additional instrument.

Limitations

The use of the SPIRR-CAD dataset for the methodological testing of the suitability of generic index instruments for capturing HRQoL of people with depressive symptoms is the key limitation. The population of patients studied had depressive symptoms as an inclusion criterion, but was simultaneously suffering from CAD, and many patients had additional comorbid illnesses. Thus, in contrast to the disease-specific instruments, the index scores of the generic instruments are not only influenced by the mental illness, but also by the limitations in the physical dimensions of the HRQoL caused by somatic comorbidity. The fact that the patients are also significantly impaired in the physical dimensions of quality of life is shown by the baseline value of 37.65 (compared to the mean value of a representative population sample standardised to 50) on the SF-36 physical health component score (PCS). This is lower than that of the SF-36 mental health component score (MCS) and also changes to a significantly lesser extent by the time of measurement after 18 months. The EQ-5D-3L may have been more influenced by the CAD and other somatic comorbidities than the SF-6D, as the former focuses predominantly on the physical dimensions, while the latter is balanced across the physical and mental dimensions.

Another limitation is associated with the value sets used. The choice of the value set for deriving the preference-based index has an impact on the result, as the preferences of the population of different countries may differ [65]. This is also very clear in this paper. The mean scores of the EQ-5D-3L$_{GER}$ and EQ-5D-3L$_{UK}$ differ greatly, while those of the EQ-5D-3L$_{UK}$ and SF-6D$_{UK}$ are relatively close. Based on these results, the decision to use the UK Value Set for hypothesis testing can be questioned. After all, German patients filled out the EQ-5D-3L and the SF-6D (via the questions of the SF-36) and the British index scores might misrepresent this self-assessment. Especially for a better comparability of the EQ-5D-3L and SF-6D (due to the fact that no German value set for the SF-6D exists) and because of the high correlation between the EQ-5D-3L$_{GER}$ and EQ-5D-3L$_{UK}$, this decision nevertheless appears to be justified.

Conclusion

Both the convergent validity and the responsiveness of the SF-6D are better than those of the EQ-5D-3L in patients with depressive symptoms. Based on the evaluated data, the SF-6D therefore appears to be more recommendable than the EQ-5D-3L for use in studies to evaluate interventions for this population. With its consistently lower standard deviation and thus more accurate estimates, the SF-6D also appears to be a more suitable instrument for cost-effectiveness studies than the EQ-5D-3L. In this regard, it would be desirable for the German context to design and conduct a valuation study for the SF-6D.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10198-022-01543-w.

Funding Open Access funding enabled and organized by Projekt DEAL. The SPIRR-CAD trial was funded by the German Research foundation (DFG; # HE 3115/10–1, HE 3115/10–2, AL 559/2–1, AL 559/2–2) and supported by the Deutsches Zentrum für Herz-Kreislaufforschung (DZHK). The Clinical Trials Center Cologne (CTC Cologne) is supported by the German Federal Ministry of Research and Education (BMBF grant 01KN1106).

Data availability statement Not applicable.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cassano, P., Fava, M.: Depression and public health: an overview. J. Psychosom. Res. **53**(4), 849–857 (2002)
2. World Health Organization: The global burden of disease. 2004 update. WHO, Geneva (2008)
3. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.L.: Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. The Lancet **367**(9524), 1747–1757 (2006)
4. Bijl, R.V., Ravelli, A.: Current and residual functional disability associated with psychopathology: findings from the Netherlands mental health survey and incidence study (NEMESIS). Psychol. Med. **30**(3), 657–668 (2000). https://doi.org/10.1017/s0033291799001841
5. Gerhards, S.A.H., Huibers, M.J.H., Theunissen, K.A.T.M., de Graaf, L.E., Widdershoven, G.A.M., Evers, S.M.A.A.: The responsiveness of quality of life utilities to change in depression: a comparison of instruments (SF-6D, EQ-5D, and DFD). Value Health **14**(5), 732–739 (2011)
6. Brazier, J., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. J. Health Econ. **21**(2), 271–292 (2002)
7. EuroQol Group: EuroQol–a new facility for the measurement of health-related quality of life. Health Policy **16**(3), 199–208 (1990)
8. Drummond, M.F.: Methods for the economic evaluation of health care programmes, 3rd edn. Oxford medical publications. Oxford Univ. Press, Oxford (2007)
9. Guyatt, G.H., Feeny, D.H., Patrick, D.L.: Measuring health-related quality of life. Ann. Intern. Med. **118**(8), 622–629 (1993). https://doi.org/10.7326/0003-4819-118-8-199304150-00009
10. Lin, X.-J., Lin, I.-M., Fan, S.-Y.: Methodological issues in measuring health-related quality of life. Tzu Chi Medi. J. **25**(1), 8–12 (2013). https://doi.org/10.1016/j.tcmj.2012.09.002
11. Khanna, D., Tsevat, J.: Health-related quality of life—an introduction. Am. J. Manag. Care **13**(Suppl 9), S218–S223 (2007)
12. Chisholm, D., Healey, A., Knapp, M.: QALYs and mental health care. Soc. Psychiatry. Psychiatr. Epidemiol. **32**(2), 68–75 (1997)
13. Luyten, J., Naci, H., Knapp, M.: Economic evaluation of mental health interventions: an introduction to cost-utility analysis. Evid. Based Ment. Health **19**(2), 49–53 (2016)
14. Korr, W.S., Ford, B.C.: Measuring quality of life in the mentally ill. Qual. Life Res. **12**(Suppl 1), 17–23 (2003)
15. Roick, C., Thierfelder, K., Heider, D., Klemm, T., Paschke, R., Angermeyer, M.C.: Untersuchung der Aussagefähigkeit psychometrischer und präferenzbasierter Lebensqualitätsindizes bei psychisch und somatisch Kranken (Quality of life instruments and health state preferences to assess effects of medical interventions for mentally and medically ill patients). Psychiatr. Prax. **31**(3), 128–137 (2004)
16. Connell, J., Brazier, J., O'Cathain, A., Lloyd-Jones, M., Paisley, S.: Quality of life of people with mental health problems: a synthesis of qualitative research. Health Qual. Life Outcomes **10**, 138 (2012)
17. Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., Jones, M.L., Paisley, S., O'Cathain, A., Barkham, M., Knapp, M., Byford, S., Gilbody, S., Parry, G.: A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess (Winchester, England) **18**(34), vii (2014)
18. Connell, J., O'Cathain, A., Brazier, J.: Measuring quality of life in mental health: are we asking the right questions? Soc. Sci. Med. **1982**(120), 12–20 (2014)
19. Richardson, J.R.J., Peacock, S.J., Hawthorne, G., Iezzi, A., Elsworth, G., Day, N.A.: Construction of the descriptive system for the assessment of quality of life AQoL-6D utility instrument. Health Qual. Life Outcomes **10**, 38 (2012)
20. Dowie, J.: Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Econ. **11**(1), 1–8 (2002)
21. Guyatt, G.: Commentary on Jack Dowie, "decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions." Health Econ **11**(1), 9–12 (2002)
22. Albus, C., Beutel, M.E., Deter, H.-C., Fritzsche, K., Hellmich, M., Jordan, J., Juenger, J., Krauth, C., Ladwig, K.-H., Michal, M., Mueck-Weymann, M., Petrowski, K., Pieske, B., Ronel, J., Soellner, W., Waller, C., Weber, C., Herrmann-Lingen, C.: A stepwise psychotherapy intervention for reducing risk in coronary artery

disease (SPIRR-CAD)–rationale and design of a multicenter, randomized trial in depressed patients with CAD. J. Psychosom. Res. **71**(4), 215–222 (2011)
23. Herrmann-Lingen, C., Beutel, M.E., Bosbach, A., Deter, H.-C., Fritzsche, K., Hellmich, M., Jordan, J., Jünger, J., Ladwig, K.-H., Michal, M., Petrowski, K., Pieske, B., Ronel, J., Söllner, W., Stöhr, A., Weber, C., de Zwaan, M., Albus, C.: A stepwise psychotherapy intervention for reducing risk in coronary artery disease (SPIRR-CAD): results of an observer-blinded, multicenter, randomized trial in depressed patients with coronary artery disease. Psychosom. Med. **78**(6), 704–715 (2016)
24. Zigmond, A.S., Snaith, R.P.: The hospital anxiety and depression scale. Acta. Psychiatr. Scand. **67**(6), 361–370 (1983)
25. Snaith, R.P.: The hospital anxiety and depression scale. Health Qual. Life Outcomes **1**, 29 (2003). https://doi.org/10.1186/1477-7525-1-29
26. Spitzer, R.L., Kroenke, K., Williams, J.B.: Validation and utility of a self-report version of PRIME-MD PHQ primary care study. JAMA **282**(18), 1737 (1999)
27. Richardson, L.P., McCauley, E., Grossman, D.C., McCarty, C.A., Richards, J., Russo, J.E., Rockhill, C., Katon, W.: Evaluation of the patient health questionnaire-9 item for detecting major depression among adolescents. Pediatrics **126**(6), 1117–1123 (2010)
28. EuroQol Research Foundation: EQ-5D-3L user guide (2018)
29. Szende, A., Oppe, M., Devlin, N.J. (eds.): EQ-5D value sets Inventory comparative review and user guide EuroQol group monographs, vol. 2. Springer, Dordrecht (2007)
30. Greiner, W., Claes, C., Busschbach, J.J.V., Graf von der Schulenburg, J.M.: Validating the EQ-5D with time trade off for the German population. Eur. J. Health. Econ. **6**(2), 124–130 (2005)
31. MVH Group: The measurement and valuation of health. Final report on the modelling of valuation tariffs. (1995)
32. Dolan, P.: Modeling valuations for EuroQol health states. Med. Care **35**(11), 1095–1108 (1997)
33. Diedenhofen, B., Musch, J.: Cocor: a comprehensive solution for the statistical comparison of correlations. PLoS ONE **10**(3), e0121945 (2015)
34. Chin, C.-L., Yao, G.: Convergent validity. In: Michalos, A.C. (ed.) Encyclopedia of quality of life and well-being research, pp. 1275–1276. Springer, The Netherlands, Dordrecht (2014)
35. Weaver, B., Wuensch, K.L.: SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. Behav. Res. Methods **45**(3), 880–895 (2013)
36. Hittner, J.B., May, K., Silver, N.C.: A Monte Carlo evaluation of tests for comparing dependent correlations. J. Gen. Psychol. **130**(2), 149–168 (2003)
37. Dunn, O.J., Clark, V.: Correlation coefficients measured on the same individuals. J. Am. Stat. Assoc. **64**(325), 366 (1969)
38. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.W.: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. **63**(7), 737–745 (2010). https://doi.org/10.1016/j.jclinepi.2010.02.006
39. Husted, J.A., Cook, R.J., Farewell, V.T., Gladman, D.D.: Methods for assessing responsiveness. J. Clin. Epidemiol. **53**(5), 459–468 (2000). https://doi.org/10.1016/S0895-4356(99)00206-1
40. Lemay, K.R., Tulloch, H.E., Pipe, A.L., Reed, J.L.: Establishing the minimal clinically important difference for the hospital anxiety and depression scale in patients with cardiovascular disease. J. Cardiopulm. Rehabilitat. Prev. **39**, E6–E11 (2019)
41. Cohen, J.: Statistical power analysis for the behavioral sciences, 2nd edn. L. Erlbaum Associates, Hillsdale (1988)
42. Kazis, L.E., Anderson, J.J., Meenan, R.F.: Effect sizes for interpreting changes in health status. Med. Care **27**(3 Suppl), S178–S189 (1989)
43. Hevey, D., McGee, H.M.: The effect size statistic: useful in health outcomes research ? J. Health Psychol. **3**(2), 163–170 (1998)
44. Martin, D.P., Engelberg, R., Agel, J., Swiontkowski, M.F.: Comparison of the musculoskeletal function assessment questionnaire with the short form-36, the Western Ontario and McMaster universities osteoarthritis index, and the sickness impact profile health-status measures. J. Bone Joint Surg. **79**(9), 1323–1335 (1997)
45. Stratford, P.W., Binkley, J.M., Riddle, D.L.: Health status measures: strategies and analytic methods for assessing change scores. Phys. Ther. **76**(10), 1109–1123 (1996)
46. Bessette, L., Sangha, O., Kuntz, K.M., Keller, R.B., Lew, R.A., Fossel, A.H., Katz, J.N.: Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Med. Care **36**(4), 491–502 (1998)
47. Angst, F., Verra, M.L., Lehmann, S., Aeschlimann, A.: Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med. Res. Methodol. **8**, 26 (2008)
48. Deyo, R.A., Centor, R.M.: Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J. Chronic Dis. **39**(11), 897–906 (1986)
49. Deyo, R.A., Diehr, P., Patrick, D.L.: Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. Control. Clin. Trials **12**(4), S142–S158 (1991)
50. Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C.W.: Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. **60**(1), 34–42 (2007)
51. Armstrong, R.A.: When to use the Bonferroni correction. Ophthalmic. physiol. optics **34**(5), 502–508 (2014). https://doi.org/10.1111/opo.12131
52. Sobocki, P., Ekman, M., Agren, H., Krakau, I., Runeson, B., Mårtensson, B., Jönsson, B.: Health-related quality of life measured with EQ-5D in patients treated for depression in primary care. Value Health **10**(2), 153–160 (2007)
53. Lamers, L.M., Bouwmans, C.A.M., van Straten, A., Donker, M.C.H., Hakkaart, L.: Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ. **15**(11), 1229–1236 (2006)
54. Supina, A.L., Johnson, J.A., Patten, S.B., Williams, J.V.A., Maxwell, C.J.: The usefulness of the EQ-5D in differentiating among persons with major depressive episode and anxiety. Qual. Life. Res. **16**(5), 749–754 (2007)
55. Gaujoux-Viala, C., Rat, A.-C., Guillemin, F., Flipo, R.-M., Fardellone, P., Bourgeois, P., Fautrel, B.: Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort. Ann. Rheum. Dis. **71**(9), 1478–1483 (2012)
56. Cameron, I.M., Crawford, J.R., Lawton, K., Reid, I.C.: Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br. J. Gen. Pract. **58**(546), 32–36 (2008)
57. Peasgood, T., Brazier, J., Papaioannou, D.: A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety. HEDS Discussion paper, 12/15. (2012). http://eprints.whiterose.ac.uk/74659/. Accessed 20 May 2022
58. Abdin, E., Chong, S.A., Seow, E., Peh, C.X., Tan, J.H., Liu, J., Hui, S.F.S., Chua, B.Y., Sim, K., Verma, S., Vaingankar, J.A., Subramaniam, M.: A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore. Psychiatry Res. **274**, 400–408 (2019)
59. Garcia-Gordillo, M.Á., Del Pozo-Cruz, B., Adsuar, J.C., Cordero-Ferrera, J.M., Abellan-Perpiñan, J.M., Sanchez-Martinez, F.I.: Validation and comparison of EQ-5D-3L and SF-6D instruments

in a Spanish Parkinson´s disease population sample. Nutr. Hosp. **32**(6), 2808–2821 (2015)
60. Marra, C.A., Woolcott, J.C., Kopec, J.A., Shojania, K., Offer, R., Brazier, J.E., Esdaile, J.M., Anis, A.H.: A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc. sci. Med. **60**(7), 1571–1582 (2005)
61. Fisk, J.D., Brown, M.G., Sketris, I.S., Metz, L.M., Murray, T.J., Stadnyk, K.J.: A comparison of health utility measures for the evaluation of multiple sclerosis treatments. J. Neurol. Neurosurg. Psychiatry **76**(1), 58–63 (2005)
62. Johnsen, L.G., Hellum, C., Nygaard, O.P., Storheim, K., Brox, J.I., Rossvoll, I., Leivseth, G., Grotle, M.: Comparison of the SF6D, the EQ5D, and the Oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet. Disord. **14**, 148 (2013)
63. Mulhern, B., Mukuria, C., Barkham, M., Knapp, M., Byford, S., Soeteman, D., Brazier, J.: Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. Brit. J. Psychiatry **205**(3), 236–243 (2014)
64. Konnopka, A., Günther, O.H., Angermeyer, M.C., König, H.-H.: Diskriminationsvermögen, Konstruktvalidität und Veränderungssensitivität des EQ-5D Lebensqualitätsfragebogens bei paranoider Schizophrenie (Discriminative ability, construct validity and sensitivity to change of the EQ-5D quality of life questionnaire in paranoid schizophrenia). Psychiatr. Prax. **33**(7), 330–336 (2006)
65. Gerlinger, C., Bamber, L., Leverkus, F., Schwenke, C., Haberland, C., Schmidt, G., Endrikat, J.: Comparing the EQ-5D-5L utility index based on value sets of different countries: impact on the interpretation of clinical study results. BMC. Res. Notes **12**(1), 18 (2019). https://doi.org/10.1186/s13104-019-4067-9

Authors and Affiliations

Maike Stolz[1,2] · Christian Albus[3] · Manfred E. Beutel[4] · Hans-Christian Deter[5] · Kurt Fritzsche[6] · Christoph Herrmann-Lingen[7,8] · Matthias Michal[4] · Katja Petrowski[9] · Joram Ronel[10,11] · Jobst-Hendrik Schultz[12] · Wolfgang Söllner[13] · Cora Weber[14] · Martina de Zwaan[15] · Christian Krauth[1,2]

1 Institute of Epidemiology Social Medicine and Health System Research, Hannover Medical School, Hanover, Germany

2 Center for Health Economics Research Hanover (CHERH), Hanover, Germany

3 Department of Psychosomatics and Psychotherapy, University of Cologne, Cologne, Germany

4 Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Mainz, Mainz, Germany

5 Department of Psychosomatics and Psychotherapy, Charité Universitaetsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany

6 Faculty of Medicine, Department of Psychosomatic Medicine and Psychotherapy, Medical Center-University of Freiburg, Freiburg, Germany

7 Department of Psychosomatic Medicine and Psychotherapy, University of Göttingen Medical Center, Göttingen, Germany

8 German Center for Cardiovascular Research (DZHK), Partner Site Göttingen, Göttingen, Germany

9 Department of Psychotherapy and Psychosomatics, Technical University of Dresden, Dresden, Germany

10 Department of Psychosomatic Medicine and Psychotherapy, Clinic Barmelweid, Barmelweid, Switzerland

11 Department of Psychosomatic Medicine and Psychotherapy, University Hospital Rechts Der Isar, Technische Universitaet München, Munich, Germany

12 Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg, Germany

13 Department of Psychosomatic Medicine and Psychotherapy, Paracelsus Medical University Nuremberg, Nuremberg, Germany

14 Department of Psychosomatic Medicine and Psychotherapy, Oberhavel Clinic, Clinic Hennigsdorf, Hennigsdorf, Germany

15 Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School, Hanover, Germany