University of Southern Denmark · Syddansk Universitet

# An Assessment of Two Generic Health-Related Quality of Life (HRQoL) Instruments in Patients Suffering from Low Back Pain

*Kim U. Wittrup-Jensen[1,3] & Jørgen Lauridsen[2]*

*(1): Bayer HealthCare AG, kim.wittrup-jensen@bayerhealthcare.com*

*(2): Institute of Public Health – Health Economics, University of Southern Denmark, jtl@sam.sdu.dk*

*(3): The study was done when Kim U. Wittrup-Jensen was a PhD student at University of Southern Denmark*

Health EconomicsPapers
2008:3

University of Southern Denmark

## *Abstract*

**Study design.** A prospective study of consecutive patients with low back pain admitted to an outpatient back pain clinic in Denmark.

**Objectives.** An empirical head-to-head comparison of the performance characteristics of two Health-Related Quality of Life (HRQoL) questionnaires, in order to assess the feasibility and validity of these two instruments in patients suffering from low back pain.

**Data material.** 296 patients with low back pain admitted to the outpatient clinic were all asked, at admission, to fill out the two generic preference-based questionnaires. Patients were given short instructions on how to fill out the questionnaires, asked to go home to complete the questionnaires, and finally to return them within enclosed pre-stamped envelopes.

**Methods.** Qualitative analysis (comparison of items), feasibility (number of missing cases per item) and features of score distribution were assessed in both the 15D and EQ-5D. Criterion validity was assessed by looking at the correlation of the (mean) score index of the 15D, EQ-5D and VAS. Construct validity, between the 15D and EQ-5D, was assessed as convergent and discriminant validity (correlation patterns and level of agreement). Further, an explanatory (common) - and a confirmatory factor analysis between the two HRQoL questionnaires were in-vestigated.

**Results**. The EQ-5D produced the lowest missing value rate. The ordinal score distribution was, for both instruments, concentrated at the upper half of the scales, indicating ceiling effects and thus reducing sensitivity and responsiveness within low back pain patients. Criterion validity was high and significant between the 15D, EQ-5D and VAS. Construct validity was fairly high between the dimensions of the 15D and EQ-5D. Differences in level of agreement were lowest between the EQ-5D profile and VAS. The explanatory factor analysis resulted in a four-factor solution with the four factors representing: (F1) a physical-motoric dimension, (F2) a mental (psychological) dimension, (F3) a senso-motoric dimension, and F4) a physical (fundamental) needs dimension. In total, the explanatory factor analysis explained approximately 52 per cent of the variance. The goodness-of-fit within the (conditional) confirmatory factor analysis was as high as 0.88, based on our *a priori* hypothesis.

**Conclusions**. A conclusive result on whether the 15D or EuroQol (including both the EQ-5D and VAS) performed uniformly as either 'best' or 'worst' in measuring HRQoL in patients with low back pain could not be obtained. Both instruments have their strengths and weaknesses. However, further research is required on how generic HRQoL instruments conform within patients with low back pain. In general, the specific features of each instrument under consideration should guide the choice of the most suitable generic HRQoL instrument in a given study.

## Introduction

A number of models have been developed for determining the values of health states at a numerical (cardinal) level of measurement [Kaplan & Anderson 1996; Gold *et al.* 1996; Kaplan 1989; Rosser *et al.* 1992; Sintonen & Pekurinen 1993; The EuroQol Group 1990; Hawthorne *et al.* 2000]. Unfortunately, different health status instruments yield different values for health states and hence different estimates of the value of health outcomes [Nord, 1996]. It is only during the last decade that researchers have begun to show an interest in comparing these different estimates and to decide which models are more valid compared to others [Gerard 1992]. According to Nord (1996), an important reason for this is that the models and instruments were viewed for a long time as tools for estimating health outcomes in terms of quality of life gained and, since there did not exist a gold standard for measuring quality of life, there was no way of judging objectively which models were more valid than others in estimating gains. As the need for prioritising of limited health care resources has become more important, the need for valid outcome estimates has increased. Hence the focus has turned to comparison of different models.

In the case of low back pain, there have been several efforts to assess preference-based generic instruments [Hurst *et al.* 1997; Suarez-Almozor *et al.* 2000; Patrick *et al.* 1995; Blake & Garrett 1997; Hollingworth *et al.* 1998; Kobelt *et al.* 1999; Wolfe & Hawley 1997]. The results are varied and inconclusive. The decision to use a generic instrument in a survey or clinical trial is often based on the nature of the research questions to be addressed, the characteristics of the population in question, the traditions of the research group, and the intellectual investments made in a given instrument used in previous research [Essink-Bot *et al.* 1997]. Relatively little attention has been given to the fact that the performance characteristics of an instrument, including feasibility and validity, may be population-specific to a greater or lesser degree. Given the increased use of generic HRQoL instruments in medical research there is a need for empirical data on the relative performance of the available generic measures among distinct patient populations.

## Objectives

The focus of the study is on the EuroQol (including the EQ-5D profile and the Visual Analogue Scale) and the 15D classification systems applied within the context of patients suffering from low back pain. Both instruments are widely used and available in many different languages. However, a review of the literature did not yield any study within the field of low back pain where the EuroQol (EQ-5D) and 15D have been compared.[1] The aim of this study is to fill this gap. The feasibility and construct validity of two multi-attribute instruments, EuroQol (including the EQ-5D profile and VAS) and the 15D, are examined. A high correlation between the two instruments is expected. Mean values are used to assess differences and agreements between the three different preference measures. The tariffs used to present the EQ-5D and 15D on a cardinal scale are based on the national Danish tariffs estimated within the general Danish population [Wittrup-Jensen *et al.* 2001; Wittrup-Jensen & Pedersen 2001].[2]

## Methods

### *Subjects*

All patients admitted to the outpatient clinic at Ringe Hospital, which is a decentralised part of Odense University Hospital, had been admitted either by their GP (77 per cent), specialists (9 per cent), chiropractor (9 per cent), or from a hospital (5 per cent).[3] Around 98 per cent of all admitted patients were living in the county of Funen. The remaining 2 per cent were living in neighbouring counties. All patients filled out a questionnaire indicating that they suffered from either specific or unspecific low back pain.

In the period from November 1999 to March 2000 all patients who were admitted to the outpatient clinic were included consecutively in the study. All 350 patients were possible candidates for inclusion in the study. 50 patients were immediately excluded, partly because they refused to participate in the study or did not have the time to wait for instructions in filling out the questionnaires, or partly because they were unable to complete the mandatory questionnaire at the preliminary examination. Of the remaining 300 patients, 4 were excluded because of severity of illness.

[1] Only one study comparing EQ-5D and 15D has been located [see Yfantopoulos and Sintonen, 2001]. However, in this study respondents were drawn from the general population. Nevertheless, their results show high similarities between the EQ-5D and the 15D.

[2] The EQ-5D tariffs are based on the parameters in the TTO3 model presented in Wittrup-Jensen *et al.* 2001.

[3] What criteria, e.g. an upper age-level, duration, severity of disease, lay behind the referral of each patient in our sample is not known. However, it is clear that patients in the sample were found to be amenable for treatment at the out-patient clinic, which may indicate that the more severe low back pain patients were not 'qualified' for referral to the clinic. This may have influenced our results, especial the features of the score distribution, and this must be acknowledged.

### *Design and material*

In total 296 patients received a questionnaire, which included the EuroQol descriptive system (including the EQ-5D profile, the VAS exercise, and the valuation task of EuroQol health states), the 15D instrument and the Low Back Pain Rating Scale. The latter, however, is not reported upon here. At admission all patients were asked to spend a few minutes looking through the questionnaire to insure that they understood the task. Patients were then asked to fill in the questionnaire at home and to return it to the outpatient clinic within fourteen days. There was no need for approval from an ethical committee. Data was processed using the statistical packages *SPSS* and *SAS* [Green *et al.* 1997; SAS Institute 1997].

### *Multi-attribute preference measures*

*EuroQol*: The EuroQol instrument is a simple, preference-based, HRQoL instrument, intended as a measure for patients receiving treatment for many different conditions [Brooks *et al.* 1991; Brooks and The EuroQol Group 1996]. The instrument has been developed by a multi-country, multidisciplinary team to provide a standardized generic instrument for both describing and valuing HRQoL [The EuroQol Group 1997]. It currently comprises a questionnaire with five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) each with three levels and a time frame of 'present day', known as the EQ-5D [Essink-Bot *et al.* 1990]. This leads to 243 ($3^5$) plausible health states plus *dead* and *unconscious.* A single index score can be estimated using information obtained from respondents filling out these five dimensions using a modelled tariff [Dolan 1997; Wittrup-Jensen *et al.* 2001].

The EuroQol instrument also includes a 'thermometer' - a Visual Analogue Scale (VAS) on which respondents are asked to rate their health status between 0 and 100, where 0 equals 'worst imaginable health state' and 100 equals 'best imaginable health state' [Glick *et al.* 1999]. Finally the instrument includes an exercise where respondents are asked to value 14 different health states and death on an analogue scale, often referred to as 'the valuation exercise' [Gudex *et al.* 1996]. The EuroQol is intended to complement other HRQoL measures and designed to be used alongside specific instruments, which may provide more detailed clinical information.

*15D*: The 15D is a preference-based instrument. It is a 15-dimensional, standardised, self-administered measure of HRQoL that can be used both as a profile and as a single index score measure for the following purposes: 1) assessment of effectiveness and efficiency (cost-utility) of health care procedures/technologies/programmes, 2) comparison of the HRQoL of population by regions/groups and over time in population studies and health surveys, 3) setting output objectives for hospitals/clinics/wards and measuring their output, 4) standardisation of patient-mix in comparing and analysing the productivity of hospitals/clinics/wards, and 5) improving clinical decision-making (a standard measure as a part of medical records) by pinpointing problems needing attention and indicating treatment results. In its original form the 15D had only twelve dimensions [Sintonen 1981]. Feedback from the medical profession led to a revision in 1986 [Sintonen 1994]. The second revision of the 15D took place in 1993 and has since then been unchanged. The health states descriptive system includes the following 15 dimensions: breathing, mental function, speech (communication), vision, mobility, usual activities, vitality, hearing, eating, elimination, sleeping, distress, discomfort and symptoms, sexual activity and depression. Each dimension is divided into five ordinal levels. The questionnaire is available in over ten languages and has been applied in a wide range of studies.

### *Analysis Plan*

*Qualitative analysis of questionnaire content*: A qualitative comparison of individual items of the 15D and EQ-5D was performed. Scales or items were considered to be comparable provided that their content was judged to refer to the same general health domain.[4]

*Feasibility*: The number of missing cases per item was assessed as an empirical indicator of feasibility. Missing values were defined as those cases in which no answer had been given, and those in which multiple responses were given when only one was required. For comparability an index was constructed accounting for the number of patients and the number of items per questionnaire.

*Features of score distribution*: The following were computed using the statistical programme *SPSS* [Green *et al.* 1997; SAS Institute 1997]: 1) number (and percentage) of patients distributed at each level within the 15D and EQ-5D; b) mean, median, range and confidence intervals on scores for 15D and the EQ-5D profile; c) a graphical distribution of scores for the 15D, the EQ-5D profile, and the VAS on a cardinal scale (0 to 1).

*Criterion validity*: This term looked at whether a (new) measure correlated with the gold standard and was assessed in two different ways.[5] First, the pattern of correlations between all three (the EQ-5D, 15D and VAS) scales, based on their cardinal scores, was examined. Since the scores on all three scales were negatively skewed (i.e. mean<median<mode), and the Kolmogorov-Smirnov test rejected the hypothesis $H_o$ that the scores were normally distributed, a nonparametric test for correlation between the three scales was applied. The Spearman rank correlation coefficient is a nonparametric version of the Pearson correlation coefficient, ranging from +1 to –1, and appropriate for interval data that do not satisfy the normality assumption. To assess criterion validity, the correlation coefficients across items in the 15D and EQ-5D profile, using the Kendall tau-*b* coefficients (also a non-parametric test), ranging from +1 to –1, based on their ordinal scores were estimated. Second, a *measurement of agreement* between the two profile indexes and the self-reported preference index used, focusing on the mean difference and limits of agreement were applied. This method was developed by Bland &

[4] This qualitative assessment was not an assessment of content validity.

Altman and is frequently recommended for method comparison studies [Altman & Bland 1983; Bland & Altman 1986].[6]

*Construct validity*: Whether the measures in question correlated with measures of other variables in hypothesised ways were assessed using convergent and discriminant validity. To assess *convergent validity,* that is does a measure of pain intensity correlate with a measure of the effects of pain, we estimated correlation coefficients for items within the 15D and EQ-5D profile. It was hypothesized that since both scales were trying to capture measurements of HRQoL, they would be highly correlated, and that those items that were conceptually related would be relatively strongly correlated, whereas those items with conceptually less in common would exhibit weaker correlations. The latter is called *discriminant validity* and focuses on whether a measure, for example of physical functioning, has a higher correlation with a measure of mental health than with a measure of mobility.

*Factor analysis:* This approach is a form of Structural Equation Modelling (SEM) and is one of the most important and powerful methods for establishing construct validity as it, compared to simple correlation analysis, attempts to provide a formal method of exploring correlation structure. Two, not different, but complementary approaches were applied. As it may not always be possible to completely explain the interrelationship between two HRQoL instruments by correlation analysis, a common factor analysis could give more information. First it was assumed that no *a priori* knowledge concerning how the dimensions of the 15D and EQ-5D profile related to each other in the case of low back pain patients was present. In order to investigate this, a common factor analysis based on the iterative *Principal Component method* followed by a *varimax rotation* of the factor pattern to look for possible higher order factors was used. This form of factor analysis is often referred to as *Exploratory Factor Analysis* (EFA).

In this approach one knows little or nothing about the factor structure. Where one has no knowledge regarding: (1) the number of factors or dimensions of excellence; (2) whether these dimensions are orthogonal or oblique; (3) the number or indicators for each factor; and (4) which dimension represents which factor [Sharma 1996], this is the method to apply. In the common factor analysis the extraction of factors is based on eigenvalues > 1, which means that the number of factors is not fixed *a priori,* but corresponds to the number of factors estimated within the analysis. The varimax rotation is an orthogonal rotation method that minimised the number of dimensions that had high loadings on each factor. It simplified the interpretation of the factors. Finally, different measures, in order to test the appropriateness of applying the factor analysis and judging how good the factor solution was, were applied.

---

[5] Since no gold standard exists, the focus is on whether the correlations between the three instruments are significant.
[6] This method will be discussed in more detailed later.

On the other hand, we did have (some) *a priori* expectations as to which dimensions were related to each other across the two HRQoL instruments, e.g. the item 'mobility'. Following these *a priori* expectations we supplemented the exploratory model with a *Confirmatory Factor Analysis* (CFA).[7] This approach assumed that the factor structure was (partly) known or hypothesized *a priori*. In other words, the complete factor structure along with the respective indicators and the nature of the pattern loadings were specified *a priori*. The objective, within the CFA, was empirically to verify or confirm the hypothesized factor structure. That was, how well did the data fit the model?

Most HRQoL instruments comprise the three general health attributes: physical, mental and social. As neither the 15D nor the EQ-5D profile contains a *social dimension*, one would expect two factors reflecting a physical and a mental health dimension, respectively. The hypothesis reflecting the *a priori* expectations within the CFA are explicitly extracted from the results revealed by the EFA.

Another reason for supplementing the EFA with a CFA is that the explanatory factor analysis suffers from at least three shortcomings. First, within a common factor analysis the factors (i.e. part of the factor solution) are assumed to be uncorrelated. However, from an intuitive point of view one would expect the abilities across factors to be somewhat interacted and thus it would be desirable to loosen the restriction of non-correlation. Second, each manifest variable (dimension) is assumed to be an indicator of each factor. From the rotated factor pattern it was expected that several of the coefficients could be restricted to zero. Third, the EFA does not cover the possibility of constructing higher-order factors. All these shortcomings may be remedied in a CFA.

## Results

### *Patient characteristics*

In total 296 patients were included in the study (cf. Table 1). 246 patients completed and returned the questionnaire, a return rate of 83 per cent. Of the patients who completed the questionnaires, 53 per cent were male and 47 per cent female. This distribution did not differ significantly from the patients who did not complete (i.e. failed to return) the questionnaire ($p$ = 0.581). Of the 246 respondents 5 had not completed the EQ-5D, the 15D and the VAS, leaving 241 cases for the analysis.

The mean age for patients included in the study was around 42 years, where the youngest patient was 15 years old and the oldest 80 years. The mean age for patients who failed to return the questionnaire was around 35 years, significantly lower ($p$ = 0.007) than that of the patients included in the study. The

[7] As suggested by Essink-Bot *et al.* (1997) more formal, confirmatory tests are needed (e.g. using structural equation models) to explore further the underlying higher-order physical and mental health score components identified in studies such as this.

main reason for this was that a relatively large group of patients ≤ 30 years of age failed to return the questionnaire and almost all of the patients ≥ 60 years completed and returned the questionnaires. Nearly 70 per cent of the patients who returned the questionnaire had a job, 8 per cent were retired and around 5 per cent were students. As in the general population, over two-thirds had continued their education beyond high school. Around 13 per cent had a university degree.

Although the return rate was fairly high compared to that normally expected in similar postal-based studied, one has to bear in mind that nearly 1/5 of the sample failed to return the questionnaires. There may be different reasons for this. One could be response burden and respondent resistance, i.e. the longer and more complex the questionnaire, the more respondents will hesitate or refuse to complete it. If questions are perceived as intrusive or too personal, respondents will also refuse to answer them. Since the questionnaire is fairly long, response burden could have played a part in respondents failing to return the questionnaires. To what degree this holds true is pure speculation.

**Table 1.** Socio-demographic characteristics of the patients in the sample.

| Variable | Completed | | Not completed | | *p*-value |
|---|---|---|---|---|---|
| | N | % | N | % | |
| *Gender:* | | | | | |
| Female | 116 | 47.2 | 24 | 48.1 | |
| Male | 130 | 52.8 | 26 | 51.9 | |
| All | 246 | 100.0 | 50 | 100.0 | 0.581[a] |
| *Age (years):* | | | | | |
| Mean | 42.1 | - | 35.5 | - | 0.007[b] |
| Median | 42.0 | - | 35.5 | - | |
| Range (min. – max.) | 15 - 80 | - | 15 - 66 | - | |
| ≤ 30 years | 46 | 18.7 | 18 | 36.0 | |
| ≥ 60 years | 23 | 9.4 | 2 | 4.0 | |
| *Occupation (n = 235):* | | | | | |
| Have job | 157 | 66.8 | na | na | |
| Retired | 19 | 8.1 | na | na | |
| Housewife | 9 | 3.8 | na | na | |
| Student | 11 | 4.7 | na | na | |
| Looking for job | 20 | 8.5 | na | na | |
| Other | 19 | 8.1 | na | na | |
| *Cont. education after high school? (n = 238):* | | | | | |
| Yes | 71 | 29.8 | na | na | |
| No | 167 | 70.2 | na | na | |
| *University degree or equivalent? (n = 231):* | | | | | |
| Yes | 200 | 85.6 | na | na | |
| No | 31 | 14.4 | na | na | |

[a] $\chi^2$-test.
[b] Independent samples *t*-test.

## *Qualitative comparison of questionnaire content*

An explicit comparison of health dimensions in the 15D and the EQ-5D profile is presented in Table 2. The *physical domain* is represented in both instruments and operationalised with emphasis on mobility. The *social role* is represented in both measurements as usual activities. The same goes for the assessment of pain where both instruments address somatic sensations other than pain by combining pain with symptoms and discomfort for the 15D and the EQ-5D profile, respectively. Both instruments have a dimension of depression, which in both cases addresses anxiety, depression or sadness. The EQ-5D profile's final dimension is self-care, which explicitly addresses the respondents'/patients' ability to wash or dress themselves, whereas the 15D has dimensions covering respondents' ability to see, hear, sleep, eat, communicate, plus their level of mental function. The 15D also includes dimensions of distress, vitality, and sexual activity.

**Table 2.** Qualitative comparison of the content of 15D and EQ-5D.

| 15D | EQ-5D |
|---|---|
| Mobility | Mobility |
| Vision | - |
| Hearing | - |
| Breathing | - |
| Sleeping | - |
| Eating | - |
| Speech (Communication) | - |
| Elimination | - |
| Usual activities | Usual activities |
| Mental function | - |
| Discomfort and Symptoms | Pain/Discomfort |
| Depression | Anxiety/Depression |
| Distress | - |
| Vitality | - |
| Sexual activity | - |
| - | Self-care |

### *Feasibility*

An overview of missing values is presented in Table 3. The EQ-5D produced the lowest number of missing values ranging from 0.4 to 2.0 per cent. The 15D showed somewhat higher, though acceptable, missing values ranging from 1.2 to 3.3 per cent. The EQ-5D also had the lowest index of 1.0. However, the index for the 15D was also at a low, and acceptable, level of 1.4.

**Table 3.** Missing values (pooled data) (n = 241).

| | Range[a] | Index[b] |
|---|---|---|
| 15D | 1.2 – 3.3 % | 1.4 |
| EQ-5D | 0.4 – 2.0 % | 1.0 |

[a]Range = range in percentage missing values per item.
[b]Index = (mean number of missing values per respondent/number of items) × 100.

### *Features of score distribution*

Table 4 illustrates how the patients scored themselves on EQ-5D. Around 55 per cent had some problems with mobility$^{EQ\text{-}5D}$; around 70 per cent had some problems with performing usual activities$^{EQ\text{-}5D}$, and almost 78 per cent were in moderate pain and/or discomfort$^{EQ\text{-}5D}$. Around 79 per cent indicated that they had no problems with self-care$^{EQ\text{-}5D}$, and around 60 per cent stated that they were not anxious/depressed$^{EQ\text{-}5D}$. Around 13 per cent were unable to perform usual activities$^{EQ\text{-}5D}$ and around 15 per cent had extreme pain/discomfort$^{EQ\text{-}5D}$.

**Table 4.** Distribution on levels in the five EQ-5D dimensions (n = 241).

| Dimension | Number of patients | Per cent |
|---|---|---|
| *Mobility:* | | |
| 1. No problems | 107 | 44.4 |
| 2. Some problems | 133 | 55.2 |
| 3. Confined to bed | 1 | 0.4 |
| *Self-care:* | | |
| 1. No problems | 191 | 79.3 |
| 2. Some problems | 50 | 20.7 |
| 3. Unable | 0 | 0.0 |
| *Usual activities:* | | |
| 1. No problems | 41 | 17.0 |
| 2. Some problems | 169 | 70.1 |
| 3. Unable | 31 | 12.9 |
| *Pain/discomfort:* | | |
| 1. No pain | 18 | 7.5 |
| 2. Moderate pain | 188 | 78.0 |
| 3. Extreme pain | 35 | 14.5 |
| *Anxiety/depression:* | | |
| 1. Not anxious | 143 | 59.3 |
| 2. Moderately anxious | 96 | 39.8 |
| 3. Extremely anxious | 2 | 0.8 |

**Table 5.** Distribution of patients on 15D dimensions and levels (per cent). (n = 241).

| *Level* / *Dimension* | No problems | Minor problems | Some problems | Moderate problems | Severe problems |
|---|---|---|---|---|---|
| Mobility | 115 (48.3) | 108 (45.4) | 13 (5.5) | 2 (0.8) | 0 (0.0) |
| Vision | 221 (92.1) | 18 (7.5) | 1 (0.4) | 0 (0.0) | 0 (0.0) |
| Hearing | 219 (90.1) | 19 (7.8) | 5 (2.1) | 0 (0.0) | 0 (0.4) |
| Breathing | 164 (68.1) | 68 (28.2) | 5 (2.1) | 3 (1.2) | 1 (0.0) |
| Sleeping | 93 (38.3) | 108 (44.4) | 34 (14.0) | 8 (3.3) | 0 (0.0) |
| Eating | 241 (99.6) | 1 (0.4) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Speech | 226 (93.0) | 14 (5.8) | 3 (1.2) | 0 (0.0) | 0 (0.0) |
| Elimination | 183 (75.3) | 52 (21.4) | 8 (3.3) | 0 (0.0) | 0 (0.0) |
| Usual activities | 40 (16.5) | 111 (45.7) | 42 (17.3) | 42 (17.3) | 8 (3.3) |
| Mental function | 200 (82.3) | 38 (15.6) | 4 (1.7) | 1 (0.4) | 0 (0.0) |
| Discomfort and Symp. | 43 (17.8) | 88 (36.4) | 87 (36.0) | 19 (7.9) | 5 (2.1) |
| Depression | 116 (48.1) | 100 (41.5) | 15 (6.2) | 8 (3.3) | 2 (0.8) |
| Distress | 110 (45.5) | 108 (44.6) | 16 (6.6) | 7 (2.9) | 1 (0.4) |
| Vitality | 67 (27.8) | 128 (53.1) | 28 (11.6) | 16 (6.6) | 2 (0.8) |
| Sexual activity | 100 (41.8) | 96 (40.2) | 30 (12.6) | 8 (3.3) | 5 (2.1) |

Note. Not all percentages equal 100 as they are only shown to one decimal place.

As illustrated in Table 5, which shows how patients scored themselves on the 15D, around 94 per cent reported no or minor problems with mobility[15D]; around 62 per cent reported no or minor problems with usual activities[15D] and around 54 per cent reported no or minor problems with discomfort and symptoms[15D].

From the same table it can also be seen that for six out of the fifteen dimensions in the 15D, over 75 per cent of the low back pain patients put themselves at the first level (no problems). For eight out of the fifteen dimensions nobody placed at level five (worst level). Not surprisingly, the majority of the low back pain patients reported 'no problems' in the dimensions vision$^{15D}$, hearing$^{15D}$, eating$^{15D}$, speech$^{15D}$, and mental function$^{15D,}$ since these dimensions were not directly related to their low back pain disease.

Table 6 illustrates the cardinal scores derived by applying weights for the EQ-5D and 15D, respectively. The VAS index represents self-reported health status on a rating scale. The 15D and the VAS are bounded by 0 (worst) to 1 (perfect health), whereas The EQ-5D allows health states worse than death, indicated by negative values. The EQ-5D is bounded by -1 (worst) to 1 (perfect health). Applying national Danish weights, the 15D resulted in a mean value of 0.85. The EQ-5D and VAS had significantly lower mean values at 0.68 and 0.62, respectively. For all three instruments the mean value was lower than the median. This indicates that there were a large number of patients with a low score. In the EQ-5D profile the fairly large difference between mean and median values can be explained by negative values. The IQ ranges differed between the three instruments. The lowest score on 15D was 0.49, whereas the lowest scores on the EQ-5D profile and VAS were -0.20 and 0.00, respectively.

**Table 6.** Summary statistics of scores using different preference measures.

| Measure | Sample median (IQ range) | Sample mean (95 % CI) | Number of cases |
|---|---|---|---|
| 15D | 0.86 (0.49 – 1.00) | 0.85 (0.84; 0.86) | 241 |
| EQ-5D | 0.71 (-0.20 – 1.00) | 0.66 (0.63; 0.69) | 241 |
| VAS | 0.65 (0.00 – 0.99) | 0.62 (0.59; 0.65) | 241 |

### *Criterion validity*

Criterion validity was assessed by the correlations between scores on (1) the 15D and EQ-5D profile, (2) the 15D and VAS, and (3) the EQ-5D profile and VAS, which were all high and significant ($p < 0.01$) (cf. Table 7). The correlation between the 15D and EQ-5D profile was higher than the correlation between both the 15D and VAS and between the EQ-5D profile and VAS.

**Table 7.** Spearman rank correlation coefficients between scores on 15D, EQ-5D profile, and VAS. (n = 241)

| | 15D | EQ-5D | VAS |
|---|---|---|---|
| 15D | 1.00 | | |
| EQ-5D | 0.71* | 1.00 | |
| VAS | 0.54* | 0.56* | 1.00 |

*($p < 0.01$).

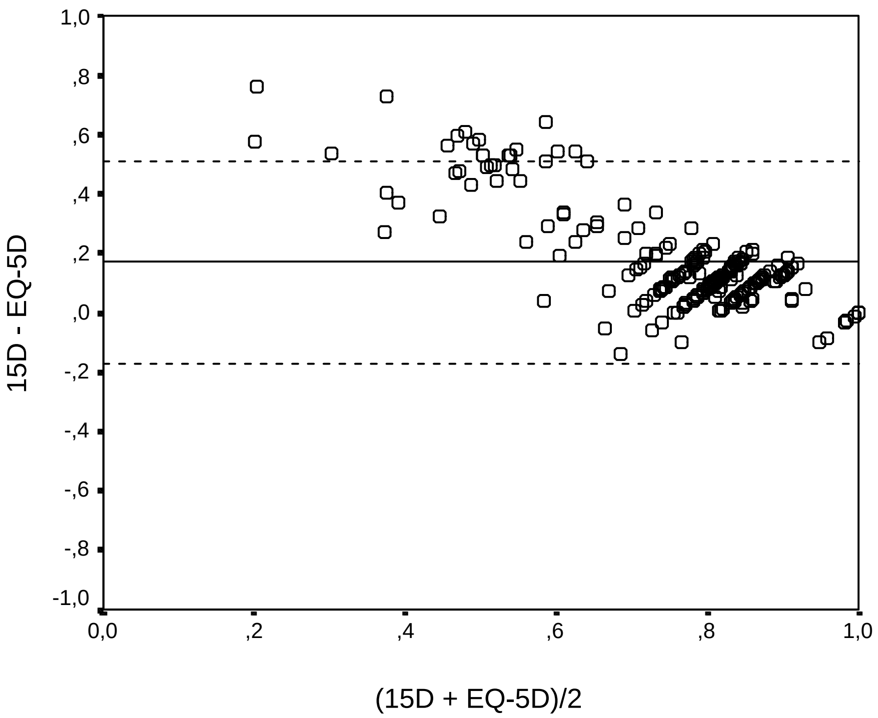
**Table 8.** Mean difference and agreement between the different preference measures. (n = 241)

| | | | | Limits of agreement (mean bias ± 2 SD) | |
|---|---|---|---|---|---|
| **Difference** | **Mean** | **95 % CI** | **SD** | **Lower** | **Upper** |
| 15D – EQ-5D | 0.17 | 0.15 to 0.19 | 0.17 | -0.17 | 0.51 |
| 15D – VAS | 0.22 | 0.20 to 0.25 | 0.18 | -0.14 | 0.56 |
| EQ-5D – VAS | 0.05 | 0.03 to 0.08 | 0.19 | -0.33 | 0.43 |

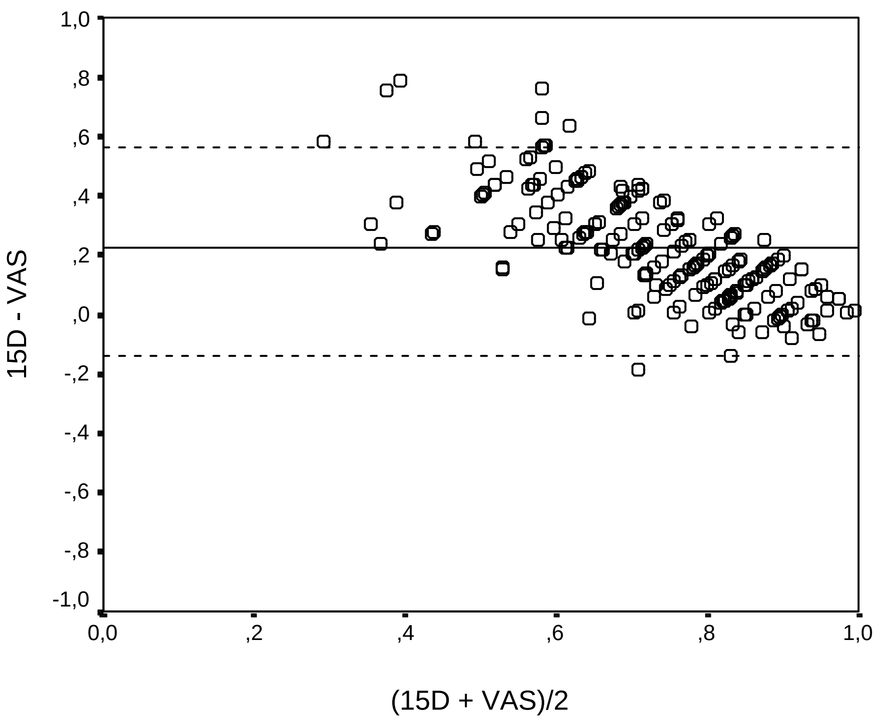
Pair-wise agreement between the 15D and both the EQ-5D profile and the VAS tended to be lower than pair-wise agreement between the EQ-5D profile and VAS, as illustrated in Table 8 for each comparison pair as the mean difference of the scores (mean bias) ± SD. In Figure 1, a plot of the pair-wise differences in scores versus the mean score for the three instruments pairs is shown for the 15D and EQ-5D profile (Figure 1a), the 15D and VAS (Figure 1b), and for the EQ-5D profile and VAS (Figure 1c). As in Table 7, figure 1c illustrates a better agreement between both Figures 1a and 1b.

**Figure 1.** Pairwise differences in scores between method pairs and means of the pairs. (a) 15D and the EQ-5D profile, (b) 15D and VAS, (c) the EQ-5D profile and VAS.

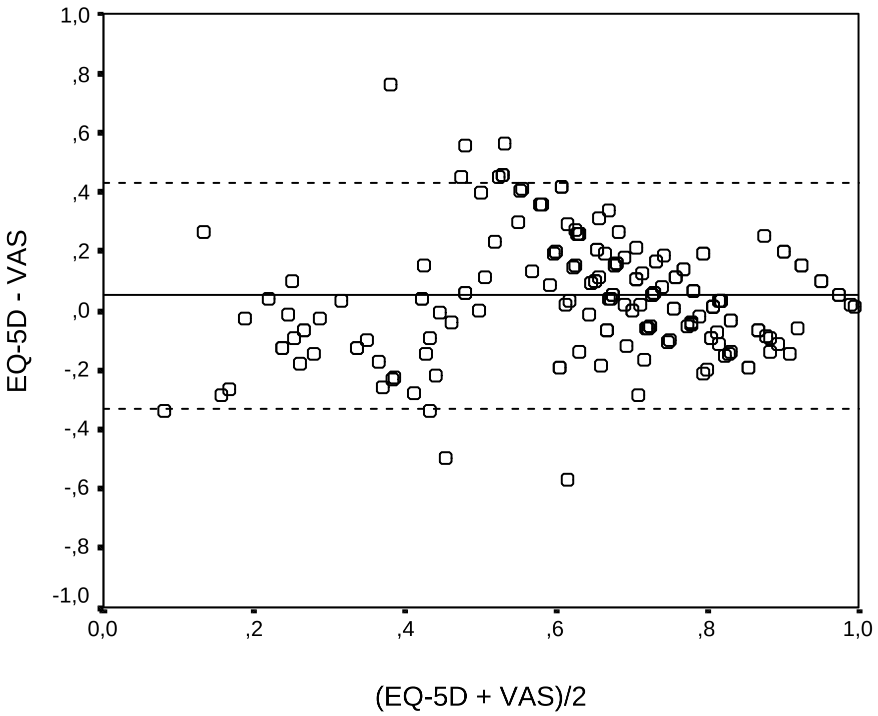
(a)




(b)



(c)




*Construct validity*

As can be seen in Table 9, all dimensions in the EQ-5D profile displayed a high and significant correlation with their identical/similar dimensions in the 15D. Patients who placed themselves high/low on the mobility$^{EQ\text{-}5D}$, usual activities$^{EQ\text{-}5D}$, and pain/discomfort$^{EQ\text{-}5D}$ dimensions also placed themselves high/low on the dimension of usual activities$^{15D}$. Further, patients who placed themselves high/low on the anxiety/depression$^{EQ\text{-}5D}$ dimension also placed themselves high/low on the distress$^{15D}$ dimension.

**Table 9.** Kendall *b*-tau correlation coefficients for the 15D and the EQ-5D profile. (n = 241).

| *15D* | *EQ-5D* | | | | |
|---|---|---|---|---|---|
| | **Mobility** | **Self care** | **Usual activities** | **Pain/discomfort** | **Anxiety/ Depression** |
| Mobility | 0.57** | 0.27** | 0.37** | 0.37** | 0.21** |
| Vision | 0.05 | -0.02 | 0.14* | -0.01 | 0.12 |
| Hearing | 0.04 | 0.07 | 0.11 | 0.02 | 0.12 |
| Breathing | 0.15* | 0.11 | 0.18** | 0.21** | 0.21** |
| Sleeping | 0.23** | 0.12 | 0.23** | 0.26** | 0.27 |
| Eating | 0.06 | 0.13* | 0.12 | 0.14* | -0.05 |
| Speech | -0.09 | -0.05 | 0.05 | 0.00 | 0.06 |
| Elimination | 0.11 | 0.09 | 0.11 | 0.24** | 0.24** |
| Usual activities | 0.49** | 0.31** | 0.65** | 0.48** | 0.29** |
| Mental function | 0.10 | 0.09 | 0.18** | 0.09 | 0.33** |
| Discomfort and Symp. | 0.39** | 0.27** | 0.45** | 0.44** | 0.24** |
| Depression | 0.16* | 0.17** | 0.36** | 0.27** | 0.68** |
| Distress | 0.05 | 0.15* | 0.20** | 0.22** | 0.64** |
| Vitality | 0.29** | 0.19** | 0.42** | 0.42** | 0.43** |
| Sexual activity | 0.33** | 0.23** | 0.41** | 0.34** | 0.36** |

**(p < 0.01).
*(p < 0.05).

Correlation within dimensions of the EQ-5D profile and 15D are illustrated in Tables 10 and 11, respectively. As can be noted, all correlation coefficients within the EQ-5D were significant at the 1 per cent level. As expected, a correlation existed between the patients' ability to move around and their ability to perform self-care$^{\text{EQ-5D}}$ and usual activities $^{\text{EQ-5D}}$. Patients who indicated that they had problems with mobility $^{\text{EQ-5D}}$ or usual activity $^{\text{EQ-5D}}$ also seemed to have problems with pain/discomfort $^{\text{EQ-5D}}$.

**Table 10.** Kendall *b*-tau correlation coefficients for EQ-5D. (n = 241)

| | **Mobility** | **Self care** | **Usual Activities** | **Pain/ discomfort** | **Anxiety/ depression** |
|---|---|---|---|---|---|
| Mobility | 1.00 | | | | |
| Self care | 0.31* | 1.00 | | | |
| Usual activities | 0.41* | 0.29* | 1.00 | | |
| Pain/discomfort | 0.43* | 0.30* | 0.44* | 1.00 | |
| Anxiety/depression | 0.18* | 0.21* | 0.32* | 0.26* | 1.00 |

*(p < 0.01).

The results were mixed for the 15D. Low back pain patients having problems with usual activities$^{\text{15D}}$ also seemed to have problems with discomfort and symptoms$^{\text{15D}}$, depression$^{\text{15D}}$, vitality$^{\text{15D}}$ and/or sexual activity$^{\text{15D}}$. Low back pain patients indicating that they had problems with depression$^{\text{15D}}$ also seemed to have problems with distress$^{\text{15D}}$, vitality$^{\text{15D}}$ and/or sexual activity$^{\text{15D}}$. There seemed to be a high and significant correlation between vision$^{\text{15D}}$ and hearing$^{\text{15D}}$, indicating that low back pain patients having problems with vision$^{\text{15D}}$ also had problems with hearing$^{\text{15D}}$. This may, however, have been a spurious effect caused by age. As expected, dimensions such as speech$^{\text{15D}}$, eating$^{\text{15D}}$, vision$^{\text{15D}}$ and/or hearing$^{\text{15D}}$ did not affect mobility$^{\text{15D}}$ or the low back pain patient's ability to perform usual activities$^{\text{15D}}$.

In addition neither vision[15D], speech[15D], eating[15D] nor hearing[15D] affected sexual activity[15D] and discomfort and symptoms[15D].

**Table 11.** Kendall *b*-tau correlation coefficients for the 15D. (n = 241).

| | MO | VIS | HE | BR | SL | EA | SP | EL | UA | MF | DS | DE | DI | VIT | SA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MO | 1.00 | | | | | | | | | | | | | | |
| VIS | 0.03 | 1.00 | | | | | | | | | | | | | |
| HE | -0.01 | 0.31[a] | 1.00 | | | | | | | | | | | | |
| BR | 0.11 | 0.11 | 0.08 | 1.00 | | | | | | | | | | | |
| SL | 0.22[a] | 0.08 | 0.14[b] | 0.20[a] | 1.00 | | | | | | | | | | |
| EA | 0.13 | -0.02 | -0.02 | -0.05 | 0.02 | 1.00 | | | | | | | | | |
| SP | 0.04 | 0.12 | 0.10 | 0.12 | 0.13[a] | -0.02 | 1.00 | | | | | | | | |
| EL | 0.11 | 0.12 | 0.20[a] | 0.17[a] | 0.24[a] | 0.15 | 0.21 | 1.00 | | | | | | | |
| UA | 0.42[a] | 0.07 | 0.05 | 0.21[a] | 0.22[a] | 0.09 | -0.02 | 0.08 | 1.00 | | | | | | |
| MF | 0.11 | 0.23 | 0.26[a] | 0.30[a] | 0.26[a] | -0.03 | 0.32[a] | 0.29[a] | 0.19[a] | 1.00 | | | | | |
| DS | 0.33[a] | 0.01 | 0.02 | 0.21[a] | 0.26[a] | 0.10 | -0.04 | 0.09 | 0.51[a] | 0.17[a] | 1.00 | | | | |
| DE | 0.20[a] | 0.11 | 0.14[b] | 0.24[a] | 0.36[a] | -0.06 | 0.21[a] | 0.19[a] | 0.33[a] | 0.33[a] | 0.25[a] | 1.00 | | | |
| DI | 0.13[b] | 0.08 | 0.14[b] | 0.22[a] | 0.23[a] | 0.05 | 0.23[a] | 0.22[a] | 0.18[a] | 0.31[a] | 0.24[a] | 0.53[a] | 1.00 | | |
| VIT | 0.26[a] | 0.06 | 0.15[b] | 0.29[a] | 0.34[a] | 0.01 | 0.14[b] | 0.27[a] | 0.46[a] | 0.34[a] | 0.39[a] | 0.50[a] | 0.30[a] | 1.00 | |
| SA | 0.31[a] | 0.07 | 0.12 | 0.17[a] | 0.34[a] | 0.03 | 0.04 | 0.23[a] | 0.40[a] | 0.23[a] | 0.43[a] | 0.46[a] | 0.28[a] | 0.44[a] | 1.00 |

[a](p < 0.01).
[b](p < 0.05).
Note: MO=Mobility, VIS=Vision, HE=Hearing, BR=Breathing, SL=Sleeping, EA=Eating, SP=Speech, EL=Elimination, UA=Usual activities, MF=Mental function, DS=Discomfort and symptoms, DE=Depression, DI=Distress, VIT=Vitality, SA=Sexual activity.

### *Results from the factor analysis*

The overall *Kaiser's Measure of sampling adequacy* (KMO) score was 0.84, which was excellent. The single-variable KMO's show that most variables (i.e. dimensions) were excellent for the analysis (see Appendix A). Exceptions were vision[15D], hearing[15D], mobility[EQ-5D], and anxiety/-depression[EQ-5D] (however, all were within the acceptable region), and eating[15D] and sexual activites[15D], which were in the questionable region. The explanatory (common) factor solution, illustrating the rotated factors, is depicted in Table 12. The rotated factor pattern comprises the coefficients for the standardized variables used to calculate the rotated factors. Thus, F1 = 0.56*Mobility[15D] + 0.05*Vision[15D] + …., where $x^* = (x - \bar{x}) / s_x$.

For each variable (dimension) that is marked with an '*', the variable has a high pattern coefficient, while a '+' indicates a fairly high coefficient. As an example, it can be seen that mobility[15D] was strongly represented in F1 and partly in F4. Looking at the column for F1, it is found that F1 strongly represented mobility[15D], usual activities[15D], discomfort and symptoms[15D], mobility[EQ-5D], self-care[EQ-5D], pain/discomfort[EQ-5D], and partly represented breathing[15D], sleeping[15D], depression[15D], vitality[15D], and anxiety/depression[EQ-5D]. Thus, F1 is considered to represent physical (motoric) ability. Similar considerations led to the following interpretations of the remaining factors: F1= *Physical-motoric dimension*, F2 = *Mental (psychological) dimension*, F3 = *Senso-motoric dimension* and F4 = *Physical (fundamental) needs dimension.*

**Table 12.** Common factor analysis of the 15D (items) and the EQ-5D profile (items).

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| *15D:* | | | | |
| Mobility | 0.56* | 0.04 | 0.03 | 0.32+ |
| Vision | 0.05 | 0.06 | 0.45* | 0.00 |
| Hearing | 0.06 | 0.09 | 0.47* | 0.02 |
| Breathing | 0.20+ | 0.26+ | 0.19 | 0.24* |
| Sleeping | 0.24+ | 0.33+ | 0.25+ | 0.18* |
| Eating | 0.12 | 0.02 | 0.01 | 0.29* |
| Speech | -0.04 | 0.20+ | 0.37* | 0.09 |
| Elimination | 0.05 | 0.20 | 0.25+ | 0.51* |
| Usual activities | 0.82* | 0.21+ | 0.09 | -0.02 |
| Mental function | 0.07 | 0.38+ | 0.57* | 0.06 |
| Discomfort and Symptoms | 0.63* | 0.22+ | 0.03 | 0.11 |
| Depression | 0.21+ | 0.79* | 0.28+ | 0.03 |
| Distress | 0.09 | 0.74* | 0.26 | 0.11 |
| Vitality | 0.44+ | 0.51* | 0.28+ | 0.13 |
| Sexual activity | 0.06 | 0.04 | -0.01 | 0.22* |
| *EQ-5D:* | | | | |
| Mobility | 0.65* | -0.02 | 0.06 | 0.22+ |
| Self-care | 0.40* | 0.07 | 0.02 | 0.19 |
| Usual activities | 0.72* | 0.21+ | 0.16 | -0.04 |
| Pain/Discomfort | 0.60* | 0.25+ | -0.05 | 0.34+ |
| Anxiety/Depression | 0.21+ | 0.74* | 0.13 | 0.08 |

Note: * indicates a high pattern coefficient + indicates a fairly high coefficient.

The fairly high coefficients for some variables (dimensions) on more than one factor indicate that the restriction of zero correlation among factors might be too restrictive. Thus, to provide an example, some of the mental ability indicators had fairly high coefficients for the physical, motoric factor.

In order to address *goodness-of-fit* of the factor solution, it can be noted that the factors (being linear combinations of the original variables (dimensions)) might be used to calculate a reconstruction or prediction of the correlation matrix [Sharma 1996]. A natural measure of goodness-of-fit is the residual correlation, i.e. the difference between the observed correlation, say $c$, and the predicted correlation, say $\hat{c}$. The Root Mean Square Residual (RMSR) is the square root of the averaged squared residual correlations. It may be calculated for the entire correlation matrix (overall RMSR) or for a single variable's correlations with the other variables. The overall RMSR was 0.04, which was very good. Likewise, the single variables (dimensions) RMSR's were fine, with a maximum as low as 0.06 for breating$^{15D}$ (see Appendix B).

Finally, it can be is noted that the commonly accepted eigenvalue-greater-than-one criterion points to a five-factor solution. However, in that solution the factors F4 and F5 merely consist of a division of the earlier F4 into two factors, where F5 is a factor for elimination$^{15D}$. Thus, though the five-factor solution is the best according to an adequacy criterion, it is not optimal with respect to a simplicity-of-the-found-solution criterion. Thus the decision was to discard it. The four-factor solution of the 15D (items) and the EQ-5D profile (items), explained around 52 per cent of the common variance, distributed on each factor as follows: F1 = 28 per cent, F2 = 12 per cent, F3 = 6 per cent, and F4 = 6 per cent.

As the explanatory (common) factor analysis (EFA) did not cover the possibility of constructing higher-order factors, we applied a confirmatory factor analysis (CFA) to test our *a priori* expectations. From Table 13, it appears obvious that the factors F1, F3, and F4 contained a common 'physical health' factor and that the F2 factor represented a 'mental health' factor. However, estimating the hypothesized model based on these *a priori* expectations alone implicitly emphasizes that the physical and mental factors are uncorrelated. The CFA revealed a goodness-of-fit of 0.69, which was low, indicating that the two dimensions were to some degree correlated with each other. Hence, a factor assuming correlation across the physical and mental factors was explicitly incorporated. In other words, the factors F1, F3, and F4 contained a common 'physical' dimension, and likewise this 'physical' factor and the 'mental (psychological)' factor may express a common factor, denoted 'general dimension'. The hypothesized model is shown below in Figure 2.

Table 13 shows a selected number of model evaluation criteria for the CFA (Appendix C gives an overview of the statistical definition of the measures).[8] Several of the goodness-of-fit indices were well over 0.8 (GFI, AGFI, CFI, RHO, DELTA2) while some of the adjusted measures were between 0.7 and 0.8 (PGFI, NFI, Rho). Only a few measures were below 0.7 (CENT, PNFI). Together, these indices provide solid evidence of a well-fitting model. It can be further noted that the RMR was very low. Finally, the $\chi^2_M$ rejected the model. However, as the $\chi^2_M$ was small compared to the baseline $\chi^2_0$ (about 1:5) we do not consider the rejection to be of practical importance.

**Figure 2.** Hypothesized model for the confimatory factor analysis.

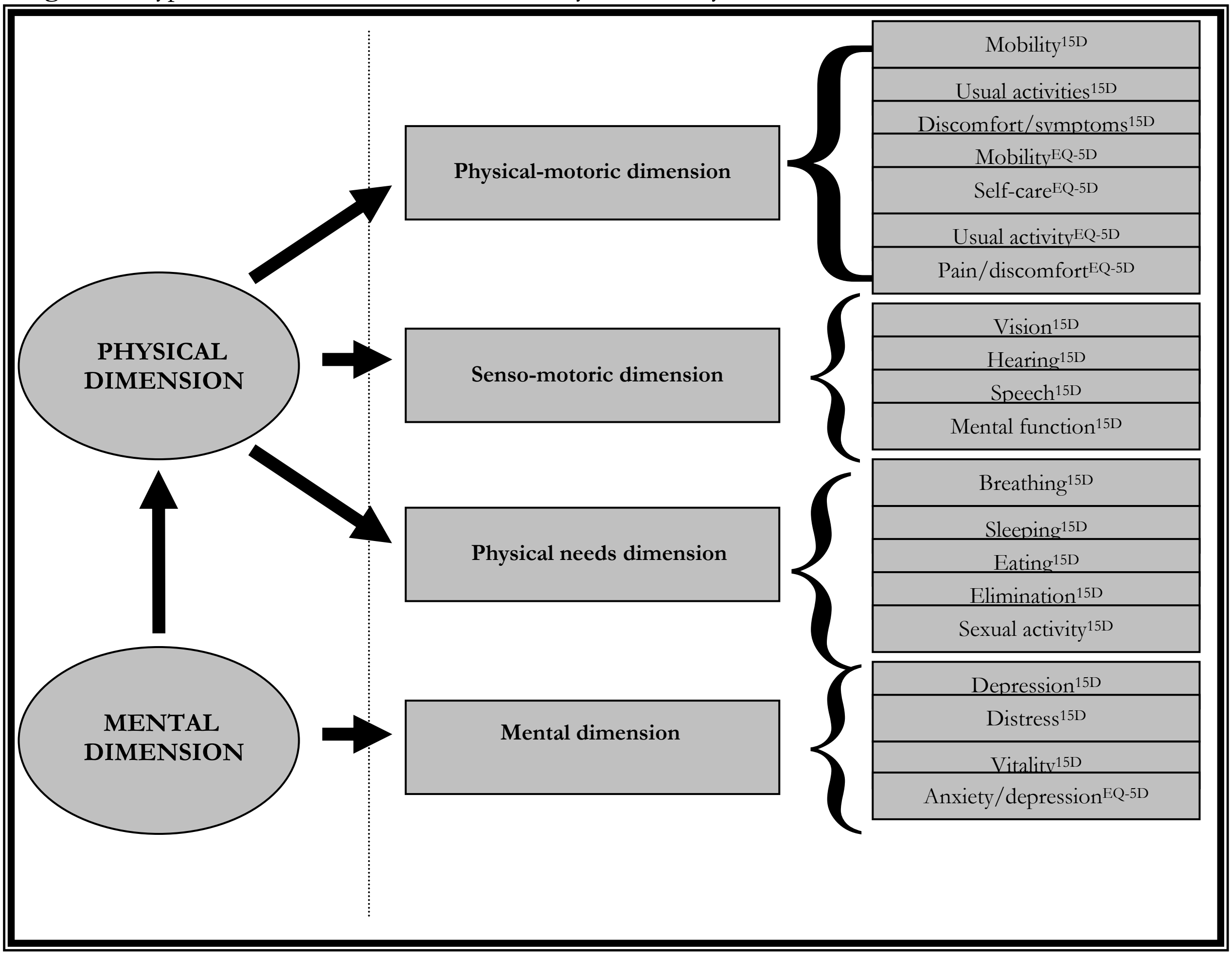



**Table 13.** Goodness-of-fit measures in the CFA model.

| Measurement | Value |
|---|---|
| Fit Function (FM) | 1.4946 |
| Goodness of Fit Index (GFI) | 0.8787 |
| GFI Adjusted for Degrees of Freedom (AGFI) | 0.8428 |
| Root Mean Square Residual (RMSR) | 0.0670 |
| Parsimonious (PGFI) (Mulaik, 1989) | 0.7492 |
| Chi-Square ($\chi^2$(M)) | 331.7904 |
| DFM | 162 |
| Pr > Chi-Square | <0.0001 |
| Independence Model Chi-Square ($\chi^2$(0)) | 1553.6 |
| DF0 | 190 |
| Bentler's Comparative Fit Index (CFI) | 0.8755 |
| Akaike's Information Criterion (AIC) | 7.7904 |
| Bozdogan's (1987) (CAIC) | 706.1714 |
| Schwarz's Bayesian Criterion (SBC) | 544.1714 |
| McDonald's (1989) Centrality (CENT) | 0.6834 |
| Bentler & Bonett's (1980) Non-normed Index (RHO) | 0.8540 |
| Bentler & Bonett's (1980) (NFI) | 0.7864 |
| James, Mulaik, & Brett (1982) Parsimonious (PNFI) | 0.6705 |
| Bollen (1986) Normed Index (Rho1) | 0.7495 |
| Bollen (1988) Non-normed Index (Delta2) | 0.8780 |

[8] For more in-depth explanation of the goodness-of-fit measures see Bearden *et al.* (1982).

All parameters were positive as expected and they all showed high significance. The estimated variances for the error terms (the specific factors) were estimated and are shown in Appendix D. The results show that the *e6* (the eating$^{15D}$ equation error term) had a high variance, thus indicating a poor fit for the 'eating' variable (dimension). The squared multiple correlations (along with the regression R-squared) were also estimated, but are not shown here. The results show that depression$^{15D}$ was well described, as F2 explained 75 per cent of its total variation. On the other hand, eating$^{15D}$ was very poor described by F4 with only 1 per cent.

Finally, the estimated model (i.e. CFA) implements a lot of restrictions, which may be omitted: (I) several of the exogenous factors (i.e. e1, …, e25, and F6 were uncorrelated, (II) the manifest variables (dimensions) were assumed to be uncorrelated with F6, and (III) several endogenous variables (dimensions) (i.e. the manifest variables and F1, …, F5) were assumed to be uncorrelated. The manifest variables (dimensions) may have been correlated with each other and with F1, …, F5. The validity of these restrictions may be examined by Lagrange Multiplier tests. We report the ten most restrictive restrictions for each of I to III in Appendix E.

## Discussion

In the present study the performance profiles of two well-known generic health status measures – the 15D and EuroQol (including the EQ-5D profile and VAS) - have been compared in a sam-ple of patients who suffer from low back pain.

Both the 15D and EQ-5D profile are single-item scales and try to address two basic health domains: physical and mental health and functioning. A qualitative comparison of these two generic HRQoL instruments indicates, however, that each instrument assesses the topic areas covered from a somewhat different perspective. Whereas the EQ-5D profile is very simple and limited in both dimensions and levels, which is exactly the idea behind the instrument when it was developed, the 15D has a much wider objective, covering dimensions such as vision, hearing, breathing, sleeping etc. From a descriptive point of view, the EQ-5D appears to cover health status among patients with low back pain quite well. However, no patients placed themselves at level 3 in the dimension 'self-care', and only 0.8 per cent placed themselves at level 3 in the dimension 'anxiety/depression'. 15D appeared to cover dimensions that are not affected by low back pain disease. In four dimensions (vision$^{15D}$, hearing$^{15D}$, eating$^{15D}$ and speech$^{15D}$) over 90 per cent of the patients placed themselves at level 1 indicating 'no problems'. Also with regard to distribution on levels, no patients were placed at level 5 in eight of the fifteen dimensions. For level 4 it was five out of fifteen. For identical/similar dimensions patients appeared to perceive these in a similar manner. With regard to the dimensions of 'mobility' and 'usual activities' the distribution across levels was almost similar. However, that was not the case for dimensions covering 'discomfort' and 'depression'. The main reason for this is that the objective is different; 'discomfort' is labelled as ‘pain/discomfort’ in the EQ-5D and as ‘discomfort and symptoms’ in the 15D. Furthermore, ‘depression’ is labelled as ‘anxiety/depression’ in the EQ-5D and as ‘depression’ in the 15D. Differences in the number of levels may also have contributed to the lack of concurrency.

The feasibility of the measures (i.e. the ease with which they can be completed by patients) was examined indirectly by calculating rates of missing values. Even though patients had most missing responses for the 15D, it is not enough to conclude that the length of an instrument has any direct bearing on the frequency of missing responses. The number of missing responses was low in both the 15D and EQ-5D profile. These findings are similar to those reported within the literature.

In one sequence we tried to apply national (Danish) tariffs in order to value patients’ health status on an interval scale ranging from 0 (worst) to 1 (perfect health) for the 15D and VAS, and from -1 (worst) to 1 (perfect health) for the EQ-5D, allowing for health states worse than death. Using the 15D resulted in a mean of 0.86. The EQ-5D resulted in 0.76 and the VAS in 0.65. These interval scores, however, are not directly comparable, which makes it hard to conclude anything about the presence/absence of criterion validity. The 15D tariffs are based on multi-attribute theory where respondents in the Danish population have valued each level in every dimension using both Category Scaling and Magnitude Estimation. The EQ-5D profile is based on a Time Trade-Off (TTO) valuation also using the Danish general population, and the VAS scores are valued directly on a Rating Scale. Furthermore, the interval differs across the three instruments as health states worse than death are allowed in the EQ-5D profile, i.e. health states can be negative. Given these disparities one should be cautious in making any direct comparisons across instruments. What can be concluded is that health status among patients with low back pain, when measured on an interval scale, did differ significantly according to which instrument was applied. The range was within 0.65 to 0.86 and the choice of instrument should be based on the given context. Nevertheless, the correlations between the instruments were statistically significant, indicating the presence of criterion validity - at least to some degree.

Adapting the suggestion put forward by Ware (1984) that a correlation coefficient of 0.40 is considered to be a substantial (construct) validity coefficient, it turned out that mobility$^{EQ\text{-}5D}$ was correlated with mobility$^{15D}$ and usual activities$^{15D}$. Usual activities$^{EQ\text{-}5D}$ was correlated with usual activities$^{15D}$, discomfort and symptoms$^{15D}$, vitality$^{15D}$ and sexual activity$^{15D}$. Pain/Discomfort$^{E!\text{-}5D}$ was correlated with usual activities$^{15D}$, discomfort and symptoms$^{15D}$ and vitality$^{15D}$. Anxiety/Depression$^{EQ\text{-}5D}$ was correlated with depresssion$^{15D}$ and distress$^{15D}$. Item specific correlations between the EQ-5D and 15D were as we expected them to be.

Using the 15D and EQ-5D profile we performed correlation analysis in order to look at both convergent and discriminant validity. As expected, and according to *a priori* findings in the literature, all dimensions within the EQ-5D profile were significantly correlated with each other at the 1 per cent level. Patients with low back pain often had problems with mobility$^{EQ\text{-}5D}$, which also caused problems with performing self-care[EQ-5D] and usual activities[EQ-5D]. Problems with mobility[EQ-5D] were high and were not unexpectedly highly correlated with pain/discomfort[EQ-5D]. More or less the same correlations were seen within the 15D instrument. As expected, patients who had problems with mobility[15D] also had problems with discomfort and symptoms[15D], depression[15D], vitality[15D] and/or sexual activity[15D]. In addition depression[15D] appeared to correlate with distress[15D], vitality[15D] and/or sexual activity[15D]. There was an unexpected correlation between hearing[15D] and vision[15D]. However, this may have been a spurious effect caused by age. A closer look into the data shows that elderly patients mainly caused this correlation, which supports our theory. Regarding discriminant validity, there was no correlation between physical dimensions such as mobility[15D] and usual activities[15D] and dimensions such as speech[15D], eating[15D], vision[15D] and/or hearing[15D], which could be expected. Further, there was no correlation between speech[15D], eating[15D], vision[15D] and/or hearing[15D] and sexual activity[15D] and discomfort and symptoms[15D].

Since it can be questioned whether applying correlation coefficients between HRQoL instruments is an appropriate technique, we also used a technique, which is referred to as *the level of agreement,* originally developed by Bland & Altman (1986). The lower the mean of pair-wise agreement, i.e. the lower the differences between the two methods, the *better*. Not surprisingly, there was a high level of agreement between the EQ-5D profile and VAS. However, this result contradicted the result we obtained from the correlation exercise, where the correlation was highest between the 15D and EQ-5D profile. The difference may be due to the fact that the two methods (i.e. correlation coefficients and level of agreement) attempt to measure different things; the correlation coefficients indicated that the mutual relationship was strongest between the 15D and EQ-5D profile, and the level of agreement indicated that to a certain degree there was an agreement between the EQ-5D profile and VAS, i.e. scores lie on or along the line of equality.

The explanatory (common) factor analysis initially resulted in a four-factor solution where F1= *Physical-motoric dimension*, F2 = *Mental (psychological) dimension*, F3 = *Senso-motoric dimension,* and F4 = *Physical (fundamental) needs dimension*. However, since we had some *a priori* expectations (hypotheses) concerning the number and meaning of the factors that a factor might represent, we applied a confirmatory factor analysis, which resulted in the factors F1, F3, and F4 containing a common 'physical' dimension, and similarly this 'physical' factor and the 'mental (psychological)' factor may express a common factor, denoted 'general dimension'. The goodness-of-fit estimates were very high for this model and provided solid evidence that the model fits our *a priori* expectations.

According to Hollingworth *et al.* (2002) it is important to establish the construct validity of all HRQoL instruments before advocating their widespread adoption. However, as noted by Brazier *et al.* (1999), establishing the validity of preference-based outcome measures is particularly difficult given the lack of an obvious gold standard. Nevertheless, such a validation process is essential in order to ensure that the valuations of the health states in question reflect patient preferences [Fitzpatrick *et al.* 1998]. In this study we assessed construct validity by using correlation analysis. Another method proposed by Brazier *et al.* (1999) is to assess the construct validity of preference-based instruments by examining the values on two hypothetical preference rules: (1) preference values should decrease as disease severity, measured by a disease-specific instrument, increases, (2) preference values in a low back pain cohort should be lower than values obtained in normative samples of the general population.

In general, the 15D and EuroQol (including the EQ-5D profile and VAS) exhibit good performance profiles in measuring HRQoL in patients with low back pain based on their construct validity. However, additional research is needed to provide a head-to-head comparison of the test-retest reliability of these instruments, as well as of other aspects of their validity, particularly including their responsiveness to changes in health status over time. Going beyond 15D and EuroQol, other generic health-related quality of life instruments may be applicable in the assessment of HRQoL in low back pain patients. Recently, Hollingworth *et al.* (2002) undertook a study of the practicality and validity of applying the SF-36 in a sample of low back pain patients and found that all the SF-36 derived preference values and directly elicited the VAS scores performed well under the two *a priori* tests of construct validity discussed above: they decreased monotonically with increasing disease severity and produced scores below that of the general population.

While feasibility (practicality) and construct validity are essential pre-requisites for any preference-based instrument, their fulfilment does not guarantee that the instrument or measure accurately reflects true preferences (for health). Furthermore, since no gold standard for measuring cardinal preferences exists, it is impossible according to Streiner & Norman (1989) to employ conventional psychometric tests of criterion validity. However, as suggested by Brazier *et al.* (1999), one could apply what they refer to as *empirical validity*, which should become the *acid test* of a preference-based instrument. Empirical validity measures the extent to which putative health state preference instruments, such as the EQ-5D, 15D, and SF-36, derive preferences that agree with the stated and revealed preferences for the health of patients. The problem would then be to elicit stated and/or revealed preferences (for health).

Finally, when choosing among available generic health status instruments it is important that the focus be not only on their formal psychometric properties but also on the match between their substantive content (e.g. the width and depth with which they address relevant health domains) and the specific research question at hand. In addition, practical considerations such as respondent burden (for example are the respondents young or old people, and how long will it take to fill out the questionnaire), and the availability of culturally- and language-adapted versions, are important in identifying the most appropriate measure for use in a given study. Another, and perhaps more important aspect when choosing an HRQoL instrument, especially if the objective is an economic evaluation, is that different instruments may yield different results, giving mixed messages concerning whether the treatment or intervention, from an economic point of view, should be implemented or not. For this reason we recommend that the use of several generic measures, or combining generic with disease-specific measures in single studies, may yield the greatest return on investment in health status assessment. Given that many of the generic HRQoL instruments are very brief and can be filled out very quickly, such a strategy should be possible without resulting in excessive response burden for the patients in question.

## References


1. Altman DG & Bland JM. Measurement in medicine: The analysis of method comparison studies. *Statistician* 1983; 32: 307-317.
2. Australian Commonwealth Department of Health, Housing and Community Services. *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee*. Canberra: Government Publication, 1992.
3. Bakker CH, Rutten-van Mölken M, van Doorslaer E, Bennett K, van der Linden S. Health related utility measurement in rheumatology: An introduction. *Patient Education and Conseling* 1993; 20: 145-152.
4. Bearden WO, Sharma S, Teel JE. Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research* 1982; 19: 425-30.
5. Blake C & Garrett M. Impact of litigation of quality of life outcomes in patients with chronic low back pain. *Irish Journal of Medical Science* 1997; 166(3): 124-126.
6. Bland JM & Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310.
7. Boyle GJ. Self-report measures of depression: Some psychometric considerations. *British Journal of Clinical Psychology* 1985; 24: 45-59.
8. Brazier J & Deverill M. A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Economics* 1999; 8: 41-51.
9. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technology Assessment* 1999; 3: 1-164.
10. Brooks RG, Jendteg S, Lindgren B, Persson U, Bjork S. EuroQol: Health-related quality of life measurement. Results of the Swedish questionnaire exercise. *Health Policy* 1991; 18: 37-48.
11. Brooks R and the EuroQol Group. EuroQol: The current state of play. *Health Policy* 1996; 37: 53-72.
12. Canadian Coordinating Office for Health Technology Assessment. *Guidelines for economic evaluation of pharmaceuticals*. 1st ed. Ottawa: Canadian Coordinating Office for Health Technology Assessment, 1994.
13. Cliff N. The eigenvalue-greater-than-one rule and the reliability of components. *Psychological Bullentin* 1988; 103(2): 276-279.
14. Comrey AL. Common methodological problems in factor analysis. *J Consult Clin Psychol* 1978; 46: 648-659.
15. Detsky AS. Guidelines for economic analysis of pharmaceutical products: A draft document for Ontario and Canada. *Pharmacoeconomics* 1993; 3: 354-63.

16. Deyo R. Measuring the Functional Status of Patients with Low Back Pain. *Archives of Physical Medicine and Rehabilitation* 1988; 69: 1044-1053.

17. Dolan P. Modeling valuations for EuroQol health states. *Medical Care* 1997; 35(11): 1095-1108.

18. Drummond MF, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford University Press. 1997.

19. Essink-Bot ML, Bonsel GJ, van der Maas PJ. Valuation of health states by the general public: Feasibility of a standardized measurement procedure. *Social Science and Medicine* 1990; 31: 1201-1206.

20. Essink-Bot ML, Krabbe PFM, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. *Medical Care* 1997; 35(5): 522-537.

21. Fairbanks J, Davies J et al. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy* 1980; 66: 271-272.

22. Fayers PM & Machin D. *Quality of life. Assessment, analysis and interpretation*. Wiley, Chichester, 2000.

23. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2: 1-74.

24. Froberg DG & Kane RL. Methodology for measuring health-state preferences – I: Measurement strategies. *Journal of Clinical Epidemiology* 1989; 42(4): 345-354.

25. Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYEs (healthy-years equi-valent) versus QALYs (quality adjusted life-years). *Journal of Health Economics* 1993; 12(3): 325-339.

26. Gerard K. Cost-utility in practice: A policy makers guide to the state of the art. *Health Policy* 1992; 21: 249-279.

27. Glick HA, Polsky D, Willke RJ, Schulman KA. A comparison of preference assessment instruments used in a clinical trial: Responses to the visual analogue scale from the EuroQol, EQ-5D and the Health Utility Index. *Medical Decision Making* 1999; 19(3): 265-275.

28. Gold M, Franks P, Erickson P. Assessing the health of the nation: The predictive validity of a preference-based measure and self-related health. *Medical Care* 1996; 34(2): 163-177.

29. Green SB, Salkind NJ, Akey TM. *Using SPSS for Windows*. Printice Hall: New Jersey. 1997.

30. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the Visual Analogue Scale. *Quality of Life Research* 1996; 5: 521-531.

31. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118: 622-629.

32. Hawthorne G, Richardson JH, Day N, Osborne R, McNeil H. *Construction and utility scaling of the assessment of Quality of Life (AQoL) instrument*. Working Paper 2000. Working Paper no. 101, Centre for Health Program Evaluation, Melbourne.

33. Hollingworth W, Dixon AK, Todd CJ, Bell MI, Antoun NM, Arafat Q, Girling S, Karia KR, Laing RJ. Self reported health status and magnetic resonance imaging findings in patients with low back pain. *European Spine Journal* 1998; 7: 369-375.
34. Hollingworth W, Deyo RA, Sullivan SD, Emerson SS, Gray DT, Jarvik JG. The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain. *Health Economics* 2002; 11(1): 71-85.
35. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: Validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology* 1997; 36: 551-559.
36. Hurst NP, Jobanputra P; Hunter M, Lambert M, Lochhead A, Brown H. Validity of EuroQol – a generic health status instrument – in patients with rheumatoid arthritis. *British Journal of Rheumatology* 1994; 33: 655-662.
37. Kaiser HF. A second generation little jiffy. *Psychometrika* 1970; 35: 401-415.
38. Kaiser HF & Rice J. Little jiffy mark IV. *Educational and Psychological Measurement* 1974; 34: 111-117.
39. Kerlinger F. *Foundations of behavioural research*. Holt, Rinehart and Winston, New York, 1986.
40. Kaplan RM. Health outcome models for policy analysis. *Health Psychology* 1989; 8: 723-35.
41. Kaplan RM & Anderson JP. The general health policy model: An integrated approach. In *Quality of life and pharmacoeconomics in clinical trials* (2nd edn), (ed. B. Spilker), pp. 309-322. Lippincott-Raven, Philadelphia, 1996.
42. Kobelt G, Eberhardt, Jönsson L, Jönsson B. Economic consequences of the progression of rheumatoid arthritis in Sweden. *Arthritis & Rheumatism* 1999; 42(2): 347-356.
43. Leaf A. Cost-effectivness as a criterion for Medicare coverage. *New England Journal of Medicine* 1989; 321: 898-904.
44. Macran S & Kind P. "Death" and the valuation of health-related quality of life. *Medical Care* 2001; 39(3): 217-227.
45. McDowell I & Newell C. *Measuring health. A guide to rating scales and questionnaires*. Oxford University Press, New York, 1996.
46. Nord E. Health status index models for use in resource allocation decisions. *International Journal of Health Technology Assessment* 1996; 12(1): 31-44.
47. Patrick DL & Erickson P. *Health status and health policy: Quality of life in health care evaluation and resource allocation*. Oxford University Press. New York, 1993.
48. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with Sciatica. *Spine* 1995; 20(17): 1899-1909.

49. Rosser R, Cotte M, Rabin R, Selai C. Index of health-related quality of life. In A. Hopkins (ed.) *Measures of the quality of life, and the uses to which they may be put*. London: Royal College of Physicians of London, 1992, 81-90.
50. SAS Institute*. SAS/STAT User's Guide Version 8*. SAS Institute: Gary, NC. 1997.
51. Sharma S. *Applied multivariate techniques*. Wiley: New York. 1996.
52. Sintonen H. An approach to measuring and valuing health states. *Social Science and Medicine* 1981; 15: 55-65.
53. Sintonen H & Pekurinen MA. A fifteen-dimensional measure of health of life (15D) and its applications. In Walker SR & Rosser RM (eds.), *Quality of life assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers, 1993, 185-195.
54. Sintonen H. *The 15-D measure of health-related quality of life: Reliability, validity and sensitivity of its health state descriptive system*. 1994. Working Paper no. 41. National Centre for Health Program Evaluation, Monash University, Melbourne.
55. Stavem K. Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of Life Research* 1999; 8: 49-54.
56. Stewart AL. Psychometric considerations in functional status instruments. In Lipkin M (eds.) *Functional status measurement in primary health care*. Springer-Verlag: New York. 1990.
57. Streiner D & Norman GR. *Health measurement scales: A practical guide to their development and use*. Oxford University Press: Oxford. 1989.
58. Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific generic and preference-based instruments. *Rheumatology* 2000; 39: 783-790.
59. The EuroQol Group, EuroQol© - a new facility for the measurement of health-related quality of life. *Health Policy 1990*; 16: 199–208.
60. The EuroQol Group. *EQ-5D User Guide*. Rotterdam: The EuroQol Group. 1997.
61. Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986; 5: 1-30.
62. Ware JE. Methodological considerations in the selection of health assessment procedures. In Wenger NK et al. (eds) *Assessment of quality of life in clinical trials of cardiovascular therapies*. Le Jacq Publishing Inc., New York, 1984.
63. Weinstein MC & Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine* 1977; 298: 716-721.
64. Wittrup-Jensen KU, Lauridsen JT, Gudex C, Brooks R, Pedersen KM. *Danish EuroQol tariffs estimated by the Visual Analogue Scale (VAS) and the Time Trade-Off (TTO).* Paper presented at the ?6th EuroQol Group Meeting in Copenhagen, 6th/7th September 2001.
65. Wittrup-Jensen KU & Pedersen KM. *Modelling weights for 15D.* Paper presented at the 22nd Nordic Study Group Meeting in Odense, 22 – 25 August 2001.

66. Wolfe F & Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. *British Journal of Rheumatology* 1997; 36: 786-793.
67. Yfantopoulos J & Sintonen H. *Comparison of the psychometric properties of the EQ-5D with 15D in Finland and Greece. A comparative analysis*. Paper presented at the EuroQol Group Meeting in Copenhagen. 2001.
68. Zwick WR & Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychological Bullentin* 1986; 99(3): 432-442.

## Appendix A

**Table A.** Kaiser's measure of sampling adequacy.

| Dimension | KMO |
|---|---|
| *15D* | |
| Mobility | 0.86 |
| Vision | 0.74 |
| Hearing | 0.79 |
| Breathing | 0.85 |
| Sleeping | 0.91 |
| Eating | 0.56 |
| Speech | 0.54 |
| Elimination | 0.81 |
| Usual activities | 0.86 |
| Mental function | 0.88 |
| Discomfort and Symptoms | 0.89 |
| Depression | 0.84 |
| Distress | 0.81 |
| Vitality | 0.91 |
| Sexual activity | 0.50 |
| *EQ-5D:* | |
| Mobility | 0.79 |
| Self-care | 0.91 |
| Usual activities | 0.89 |
| Pain/Discomfort | 0.91 |
| Anxiety/Depression | 0.76 |

## Appendix B

**Table B.** Root mean square off-diagonal residuals (RMSR).

| Dimension | RMSM |
|---|---|
| *15D* | |
| Mobility | 0.05 |
| Vision | 0.04 |
| Hearing | 0.04 |
| Breathing | 0.06 |
| Sleeping | 0.04 |
| Eating | 0.05 |
| Speech | 0.05 |
| Elimination | 0.03 |
| Usual activities | 0.03 |
| Mental function | 0.03 |
| Discomfort and Symptoms | 0.03 |
| Depression | 0.03 |
| Distress | 0.04 |
| Vitality | 0.04 |
| Sexual activity | 0.05 |
| *EQ-5D:* | |
| Mobility | 0.04 |
| Self-care | 0.04 |
| Usual activities | 0.03 |
| Pain/Discomfort | 0.03 |
| Anxiety/Depression | 0.05 |

## Appendix C

Using Maximum Likelihood estimation of the CFA model, the fit function to be minimized with respect to $\Sigma$ is

$$F_M = \log(\det(\Sigma)/\det(S)) + tr(\Sigma^{-1}S) - q \qquad (C.1)$$

where *S* is the observed covariance matrix for the q manifest variables and $\Sigma$ the estimate of *S* from the CFA. Model evaluation criteria may be divided into two groups. *Absolute measures* compare $\Sigma$ to *S*, with some correction for degrees of freedom and/or sample size (as for the regression $R^2$ or other GFI measures). *Relative measures* compare $\Sigma$ or $F_M$ to the simplest possible baseline model estimates $\Sigma$ or $F_0$ (as with the regression F test which compares a model to the intercept-only model).

*Absolute measures* employed are the following:

The goodness-of-fit index (GFI) measures the proportion of variance and covariance explained:

$$GFI = 1 - tr[(\Sigma^{-1}S - I)^2]/tr[(\Sigma^{-1}S)^2] \qquad (C.2)$$

which resembles the regression $R^2$ value. Letting $s_1$ be the number of distinct variances and covariances in *S* (thus, $s_1 = q(q+1)/2$) and $s_0$ the number of parameters estimated to calculate $\Sigma$, the degrees of freedom of the model is $df_M = s_1 - s_0$. The GFI adjusted for degrees of freedom (AGFI) reads as

$$AGFI = 1 - (s_1/df_M)(1-GFI) \qquad (C.3)$$

and penalizes models which have a high GFI at the cost of many parameters. RMR is the square root of the squared residuals,

$$RMR = \sqrt{[(2/(n(n+1))\ \Sigma_{i=1..N}\Sigma_{j=1..i}\ (s_{ij} - c_{ij})^2]} \qquad (C.4)$$

where $s_{ij}$ and $c_{ij}$ refer to elements ij from S and $\Sigma$. The $\chi^2$ is a $\chi^2$ distributed test for $H_0$: $\Sigma = S$, calculated as

$$\chi^2 = N^*F_M , \qquad (C.5)$$

with $df_M$ degrees of freedom. It is generally accepted that the test is not very reliable due to the dependence on *N*, leading to a strong tendency to reject even very well-fitting models (see Bearden, Sharma and Teel 1982). Akaikes Information Criterion is calculated as

$$AIC = \chi^2 - 2df_M \quad (C.6)$$

thus adjusting the $\chi^2$ for degrees of freedom. Adjustments for sample size are accomplished by the Bozdogan consistent AIC,

$$CAIC = \chi^2 - (\ln(n)+1)df_M \, , \quad (C.7)$$

the Schwarz Bayesian Criterion

$$SBC = \chi^2 - \ln(N)df_M \, , \quad (C.8)$$

and the McDonald measure of centrality

$$CENT = \exp(-(\chi^2 - df_M)/(2N)). \quad (C.9)$$

*Relative measures* compare the estimated $\Sigma$ to a simplest possible estimate $\Sigma_0$, rather than to the observed *S*. This baseline model suggests that no factors underlie the observed variables and that the correlations between the observed variables are zero.

Assuming that $F_0$ is the fit function calculated with $\Sigma_0$, the baseline $\chi_0^2$ is calculated as

$$\chi_0^2 = NF_0 \, . \quad (C.10)$$

The $\chi_0^2$ can serve as a benchmark while interpreting the $\chi^2$ of the less restricted model, thus indicating the increase in fit obtained by this model. The parsimonious Normed Fit Index measures the relative improvement in the fit function, adjusting for parsimony of the model (understood as a high $df_M$):

$$PNFI = (df_M/df_0)(F_0-F_M)/F_0 \, . \quad (C.11)$$

The Bentler-Bonett non-normal index RHO is defined as

$$RHO = (F_0/df_0 - F_M/df_M)/(F_0/df_0 - 1/N) \quad (C.12)$$

and measures the relative improvement in the fit function with a finite-sample and a degrees-of-freedom correction. The Butler-Bonett NFI is the unadjusted version,

$$NFI = (F_0 - F_M)/F_0 \, . \quad (C.13)$$

The normal index Rho1 resembles the PNFI, as

$$\mathrm{Rho1} = (F_0/df_0 - F_M/df_M)/(F_0/df_0) \quad \text{(C.14)}$$

while the non-normed index Delta2 is defined as a modification of the NFI reading as

$$\mathrm{Delta2} = (F_0 - F_M)/(F_0 - df_M/N) \quad \text{(C.15)}$$

## Appendix D

**Table D.** Variance of the specific factors.

| Parameter | Estimate | Standard error | *t*-value | *R*-Square |
|---|---|---|---|---|
| e1 (mobility$^{15D}$) | 0.66468 | 0.06803 | 9.77 | 0.3353 |
| e2 (vision$^{15D}$) | 0.88124 | 0.08819 | 9.99 | 0.1188 |
| e3 (hearing$^{15D}$) | 0.85032 | 0.08660 | 9.82 | 0.1496 |
| e4 (breathing$^{15D}$) | 0.75665 | 0.08402 | 9.01 | 0.2434 |
| e5 (sleeping$^{15D}$) | 0.65738 | 0.08373 | 7.85 | 0.3426 |
| e6 (eating$^{15D}$) | 0.98534 | 0.09409 | 10.47 | 0.0147 |
| e7 (speech$^{15D}$) | 0.79850 | 0.08422 | 9.48 | 0.2015 |
| e8 (elimination$^{15D}$) | 0.81286 | 0.08578 | 9.48 | 0.1872 |
| e9 (usual activities$^{15D}$) | 0.30364 | 0.04332 | 7.01 | 0.6964 |
| e10 (mental function$^{15D}$) | 0.39932 | 0.09321 | 4.28 | 0.6007 |
| e11 (discomfort and symptoms$^{15D}$) | 0.54319 | 0.05872 | 9.25 | 0.4568 |
| e12 (depression$^{15D}$) | 0.24593 | 0.03919 | 6.28 | 0.7541 |
| e13 (distress$^{15D}$) | 0.43102 | 0.04954 | 8.70 | 0.5690 |
| e14 (vitality$^{15D}$) | 0.53455 | 0.05720 | 9.34 | 0.4655 |
| e15 (sexual activity$^{15D}$) | 0.97780 | 0.09368 | 10.44 | 0.0222 |
| e16 (mobility$^{EQ\text{-}5D}$) | 0.59887 | 0.06293 | 9.52 | 0.4011 |
| e17 (self-care$^{EQ\text{-}5D}$) | 0.81332 | 0.07982 | 10.19 | 0.1867 |
| e18 (usual activity$^{EQ\text{-}5D}$) | 0.42861 | 0.05058 | 8.47 | 0.5714 |
| e19 (pain/discomfort$^{EQ\text{-}5D}$) | 0.54853 | 0.05912 | 9.28 | 0.4515 |
| e20 (anxiety/depression$^{EQ\text{-}5D}$) | 0.36940 | 0.04544 | 8.13 | 0.6306 |
| e21 (F1) | 0.29014 | 0.03657 | 7.93 | 0.3200 |
| e22 (F3) | 0.07878 | 0.02108 | 3.74 | 0.5148 |
| e23 (F4) | 0.09865 | 0.06009 | 1.64 | 0.7549 |
| e24 (F2) | 1.00000 | - | - | 0.8632 |
| e25 (F5) | 1.00000 | - | - | 0.9233 |

## Appendix E

**E.I.** Cov(ei, ej) = 0 for $i \neq j$, i, j = 1, …, 25:

| Row | Column | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| e24 (F2) | e14 (vitality$^{15D}$) | 31.42 | < 0.0001 |
| e25 (F5) | e14 (vitality$^{15D}$) | 31.42 | < 0.0001 |
| e14 (vitality$^{15D}$) | F6 | 31.41 | < 0.0001 |
| e21 (F1) | e14 (vitality$^{15D}$) | 28.69 | < 0.0001 |
| e16 (mobility$^{EQ\text{-}5D}$) | e1 (mobility$^{15D}$) | 22.02 | < 0.0001 |
| e18 (usual activity$^{15D}$) | e9 (usual activities$^{15D}$) | 16.31 | < 0.0001 |
| e20 (anxiety/depression$^{EQ\text{-}5D}$) | e13 (distress$^{15D}$) | 15.65 | < 0.0001 |
| e20 (anxiety/depression$^{EQ\text{-}5D}$) | e7 (speech$^{15D}$) | 15.63 | < 0.0001 |
| e15 (sexual activity$^{15D}$) | e4 (breathing$^{15D}$) | 14.29 | 0.0002 |
| e3 (hearing$^{15D}$) | e2 (vision$^{15D}$) | 11.90 | 0.0006 |

**E.II.** The manifest variables (dimensions) are assumed to be uncorrelated with F6:

| Row | Column | Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Vitality$^{15D}$ | F6 | 31.41 | <0.0001 |
| Depression$^{15D}$ | F6 | 11.26 | 0.0008 |
| F4 | F6 | 8.10 | 0.0044 |
| Mobility$^{EQ\text{-}5D}$ | F6 | 6.09 | 0.0136 |
| F3 | F6 | 3.64 | 0.0565 |
| Sexual activity$^{15D}$ | F6 | 3.35 | 0.0672 |
| Sleeping$^{15D}$ | F6 | 3.01 | 0.0830 |
| Pain/discomfort$^{EQ\text{-}5D}$ | F6 | 1.76 | 0.1844 |
| Breathing$^{15D}$ | F6 | 1.18 | 0.2781 |
| Usual activity$^{EQ\text{-}5D}$ | F6 | 1.09 | 0.2960 |

**E.III.** Several endogenous variables (dimensions) (i.e. the manifest variables and F1,…, F5) are assumed to be uncorrelated.

| Row | Column | Chi-Square | P > ChiSq |
|---|---|---|---|
| Vitality$^{15D}$ | F1 | 33.80 | < 0.0001 |
| F2 | Vitality$^{15D}$ | 31.43 | < 0.0001 |
| F5 | Vitality$^{15D}$ | 31.42 | < 0.0001 |
| Vitality$^{15D}$ | F5 | 31.41 | < 0.0001 |
| F1 | Vitality$^{15D}$ | 28.43 | < 0.0001 |
| Vitality$^{15D}$ | Usual activities$^{15D}$ | 26.31 | < 0.0001 |
| Vitality$^{15D}$ | Pain/discomfort$^{EQ\text{-}5D}$ | 22.60 | < 0.0001 |
| Vitality$^{15D}$ | F4 | 22.47 | < 0.0001 |
| Mobility$^{15D}$ | Mobility$^{EQ\text{-}5D}$ | 22.02 | < 0.0001 |
| Mobility$^{EQ\text{-}5D}$ | Mobility$^{15D}$ | 22.02 | < 0.0001 |