Psychology Science Quarterly, Volume 50, 2008 (1), p. 7-20

Factorial validity of the Short Form 12 (SF-12) in patients with diabetes mellitus

CARSTEN MAURISCHAT[1,2], PETER HERSCHBACH[3], ACHIM PETERS[4] & MONIKA BULLINGER[5]

Abstract

Measuring quality of life is considered as an important outcome criterion in clinical studies. Generic instruments as the Short Form 12 Health Survey offer the possibility to compare outcomes among different indications. Therefore a test of the factorial (structural) validity of the method in each indication is necessary. This study is based upon SF-12 data from 343 patients with diabetes mellitus, which were pooled together from two rehabilitation research projects. The psychometric properties of the measure were analyzed, and the questionnaires structure was tested using confirmatory structural equation modeling. In a second step age and gender specific analyses were undertaken.

The questionnaires psychometric properties prove to be comparable to international results. The structural analyses support a model that specifies covariations of errors between two items of same wording („accomplished less"). SF-12 physical and mental (latent) components are highly associated. Therefore construct validity must be criticized. Moreover both physical and mental health seem not to be independent in patients view, since items response is influenced by same wording more than subjective health reception. SF-12 summary scores should not be seen and misinterpreted as independently.

Key words: quality of life, SF-12, Diabetes

[1] Corresponding author: Carsten Maurischat, PhD, Leibniz-Institute for Science Education, Olshausenstr. 62, D-24098 Kiel; phone: +49 431 880 3097, fax: +49 431 880 5242, email: c.maurischat@gmx.de

[2] Carsten Maurischat, PhD; Institute and Policlinic of Medical Psychology, University Hospital Hamburg-Eppendorf, Germany

[3] Peter Herschbach, MD, Professor; Institute of Psychosomatic Medicine, Medical Psychology, and Psychotherapy, Technical University Munich, Germany

[4] Achim Peters, MD, Professor; Medical Clinic I, Medical University of Luebeck, Germany

[5] Monika Bullinger, PhD, Professor; Institute and Policlinic of Medical Psychology, University Hospital Hamburg-Eppendorf, Germany

Introduction

Measuring of quality of life as a representation of subjective health from the patient's point of view respectively is highly relevant for health care (Kaplan, 2003) and specifically within diabetes type 1 and 2 (Beaser et al., 1996). The main goals in diabetes treatment beside the adjustment and control of blood-glucose level are reduction of symptoms, prevention of co-morbidities, maintenance of physical functions, and maximization of quality of life. Complications, for example amputations, are related to impairments in social functioning and quality of life (Price & Harding, 2000). Thus quality of life is an important outcome of intervention studies (Norris et al., 2001; Wiesinger et al., 2001), or in epidemiological cross-section research (Camacho et al., 2002).

Up to date it is unclear, whether disease specific (targeted) instruments, such as the Diabetes Quality of Life Measure (DQOL; DCCT, 1988) or the Diabetic-specific Quality of Life Scale (DSQOLS; Bott et al., 1998) among others (Garrat et al., 2002) are to be preferred to generic, disease comprehensive instruments, such as the Short Form 36 (SF-36) (Ware, 2000). The Short Form 36 as well as its shortened version, the Short Form 12, is recommending to be used as a generic instrument in patients with diabetes (Johnson & Coons, 1998). In comparison with a diabetes-specific instrument the Short Form shows good reliability and construct validity, but is also susceptible to existing co-morbidities (Woodcock et al., 2001). Therefore, these authors recommend the use of a disease specific instrument and the generic Short Form simultaneously.

When using generic instruments in new populations and indications respectively, it is necessary to examine the basic assumptions of scale construction, especially the factorial validity. This proof is essential for a generic instrument in order to allow the comparison of results that were obtained through other indications (Anderson et al., 1996; Reed, 1998). While several studies exist on the long form SF-36, there are only few investigations about its short version Short Form 12 (SF-12). This lack of research onto SF-12 is surprising because the items of the SF-12 are part of the SF-36, and extracting this short version would be simple. For example Johnson and Maddigan (2004) examine the validity of the Short Form standard algorithm for the calculation of the weighted summary scales in diabetes patients. They developed an alternative algorithm, which is used increasingly in clinical research. However, the examination of the factorial validity of the SF-12 questionnaire in patients with diabetes has not been accomplished yet.

The objective of this study was to reproduce the psychometric properties and the scale structure of the SF-12 in a sample of patients with diabetes mellitus, and to compare the results with international findings (Amir et al., 2002; Gandhi et al., 2001; Jenkinson & Layte, 1997; Salyers et al., 2000). Using structural equation modeling (SEM) the scale structure was also proven confirmatorily, whereas the theoretical assumptions were derived from international findings (Wilson et al., 2002). Additional analyses concerned possible differences in age and gender on the items and summary scales in this sample.

Materials and methods

Design of the study

We used pooled data for this secondary data analysis. Within the framework of a national network cooperation in rehabilitation sciences, data of the SF-12 as well as patients gender, and age were recorded from two research projects. Data were combined into one common data set. In the first study the SF-36 was administered to the patients. For our analysis the SF-12 items were extracted from the SF-36 long version. According to Hurst et al. (1998) no differences are to be expected using this proceeding compared to a priori use of SF-12. In the second project the SF-12 itself was administered to patients at the beginning of a rehabilitation intervention program.

Instrument

The SF-12 is the short version of the Short form 36 Health Survey which was developed within the framework of the Medical Outcome Study (MOS) (Ware et al., 1996). Using 12 items the SF-12 assesses the two main dimensions of quality of life: physical and mental health. These dimensions reflect eight sub-dimensions: physical functioning (PF, two items), role physical (RP, two items), bodily pain (BP, one item), general health (GH, one item), vitality (VT, one item), social functioning (SF, one item), role-emotional (RE, two items), and mental health (MH, two items) (see table 1). The items show different scale levels: two,

Table 1:
Items and factor loadings of the SF-12

			physical component		mental component	
no.	**Items**	**subdimension**	**A§**	**B #**	**A§**	**B #**
2	moderate activities	Physical Functioning (PF)	.71 - .86	.61	.04 - .27	.20
3	climb several flights		.69 - .78	.70	-.03 - .13	.15
4	accomplished less	Role Physical (RP)	.68 - .74	.60	.22 - .33	.52
5	limited in kind		.62 - .76	.71	.13 - .36	.35
8	pain interfere	Bodily Pain (BP)	.67 - .74	.77	.21 - .33	.31
1	rating of general health	General Health (GH)	.41 - .72	.61	.21 - .46	.38
10	energy	Vitality (VT)	.30 - .54	.33	.42 - .58	.68
12	frequency health problems interfered	Social Functioning (SF)	.28 - .61	.35	.44 - .71	.69
6	accomplished less	Role Emotional (RE)	.16 -.24	.18	.72 - .78	.74
7	not careful		.15 - .20	.11	.69 - .73	.75
9	peaceful	Mental Health (MH)	.11 - .20	.08	.67 - .75	.77
11	downhearted/low		.04 - .10	.16	.72 - .78	.75

§ comparison values according to (Amir et al., 2002; Gandhi et al., 2001; Jenkinson & Layte, 1997)
# results in this sample

three, five or six answer categories. According to the theoretical test model, the first four sub-dimensions and their six items are indicators of the physical health component (PC), and the six items of the last four sub-dimensions are indicators of the mental health component (MC), respectively. Using item specific weighted indicators, summary scores for both physical (PSS) and mental health (MSS) could be calculated. The transformed SF-12 scores are standardized from 0 to 100. Higher values represent a higher quality of life. In contrast to the SF-36, the calculation of eight sub-dimensions scores is not possible using the SF-12.

Due to the algorithm all 12 items need to be completed in order to calculate the summary scores of physical and mental health. With the SF-12 this often leads to missing values in the main dimensions (Lim & Fisher, 1999).

Statistical analyses

Questionnaires structure and psychometric properties

Exploratory Factor Analyses (EFA)

To analyze the structure of the questionnaire, two principal component factor analyses (PCA) were calculated. The aim of the EFA was to compare the results with international findings. So the methodological approach was undertaken according to these previous studies. A first (non-rotated) factor analysis was calculated to determine the number of main latent components according to the scree plot (criteria: eigen values greater 1). In order to ensure the comparability with other studies mentioned above, and the Short Form manual the second PCA was forced to extract two components and we used Varimax-Rotation to improve the simple structure of the loading matrix. According to Ware et al. (1996) it can be assumed that the six items indicating physical health show their highest loadings on one of both components (the “physical component”) while the six remaining items that assess mental health show maximal loadings on the second (“mental”) component. However, cross-loadings on the other than theoretically assumed component are often found in psychometric Short Form literature in conjunction with the items of the sub-dimensions GH, VT, and SF. Here, cross-loadings are viewed as substantial when they are associated $a_{ij} > .40$ (Garratt et al., 2002).

Confirmatory Factor Analyses (CFA)

In contrast to exploratory factor analysis structural equation modeling makes it possible to test the hypothetical structure of a questionnaire confirmatively and to estimate the model’s validity for given empirical data. In the current analyses two models were tested that have shown empirical evidence. Therefore the theoretical assumptions tested here, are not based on the results of the EFA mentioned above, but on the empirical findings of other studies. First, a theoretical model assuming two latent correlated factors without cross-loadings was tested in which the items 1, 2, 3, 4, 5, and 8 are indicators of physical health and the items 6, 7, 9, 10, 11, and 12 are indicators of mental health (cf. table 1). Based on the results of Wilson et al. (2002), a second model was specified and covariations between the error of the items that belong to the same sub-dimensions (e.g. between item 2 and 3, which are both indicator items of “physical functioning”) were allowed. All confirmatory analyses were calculated with AMOS 4.0 (http://www.smallwaters.com/amos/). In order to estimate the parameters of the model different methods exist that mostly require interval scaled, multivariate normal distributed data (Ogasawara, 2003), newer simulation studies using various samples survey sizes and sample survey distributions show that the violation of the normal distribution assumption alone has only a minimal influence on the performance of the models fit, if the models are specified correctly (Hu & Bentler, 1998). According to Wilson et al. (2002) a maximum-likelihood-estimation of the correlation matrix was conducted, even though strictly speaking no interval scaled items are available. Alternative methods, for example distribution-free methods (ADF, asymptotic distribution free), require substantially larger samples of at least 2500 persons, so this method could not be used.

The model's fit was determined using different fit indices (Hu & Bentler, 1999; Hu et al., 1992; Hurley et al., 1997). The relationship of the χ^2-value and the degrees of freedom was supposed to be χ^2/df < 6. The GFI (goodness of fit index) as an index of the variance explained by the model, as well as the AGFI (adjusted goodness of fit index), should be in an acceptable range of (A)GFI > .89. The same criterion was set for the incremental indices like the NFI (normed fit index), the TLI (Tucker-Lewis-Index), and the CFI (comparative fit index). As a measure of parsimony values for the PGFI (parsimony goodness of fit index) and the PNFI (parsimony normed fit index) PGFI, PGNI > .39 are expected. SRMR (standardized root mean square residual) and the RMSEA (root mean square error of approximation) as functions of the differences between the estimated population covariance matrix and the sample covariance matrix are viewed below SRMR, RMSEA < .11 as acceptable and SRMR, RMSEA < .06 as good. The intercorrelations of the items of the SF-12 and the correlations of these items with the summary scales as well as age and gender were calculated using Spearman-Correlation.

Results

Sample survey

The total survey sample of the two projects includes 366 patients. After the calculation of the summary scales 343 patients remain for the pooled analysis (6.3 % missing values). About 82 % of the sample is type 2 diabetes mellitus patients, and 18 % have type 1 diabetes mellitus. The sample included 108 women (31.5 %) and 235 men (68.5 %) with an average age of 51.9 years (SD = 7.8, range: 23 to 76). Due to the small proportion (< 100 cases) of type 1 diabetes mellitus patients, psychometric analyses for this subgroup separately could not be done with trust. Furthermore because of the generic alignment of the SF-12`s factorial validity, it is justified to analyze both kinds of patients together.

Structure of the questionnaire and psychometrics properties

The scree plot of the eigen values (6.1; 1.2; 0.9; 0.6; 0.6) might suggest a two-factorial solution. The following PCA for two components using Varimax-Rotation explains 60.3 % variance (KMO = .90; Bartlett Test: χ^2 = 2160.6; df = 66, p = .000). Factor loadings of the 12 items onto the components can be found in table 1 (column B) in which comparison results from Salyers et al. (2000), Gandhi et al. (2001), and Amir et al. (2002) are used (see column

A of table 1). It shows that in this sample survey the factor loadings of all items were highest on each component assumed. According to Ware (2000) the loading of items on the second component is $a_{ij} < .40$ and therefore without substantial significance. Exception is the item role physical (RP) no. 4 "less accomplished" that shows a very high crossloading on the mental component beyond the range of the comparative studies presented in column A.

Confirmatory factor analysis

The distribution of the 12 items shows values of under 1 and no major deviation in skewness and kurtosis from an approximately standard normal distribution. The results in terms of goodness of fit indices of the two specified models (according to Wilson et al., 2002) can be found in table 2 (cf. model 1 and model 2). Inspirited by the results of our exploratory analysis presented before, a third (nested) model was created with an additional covariation between the measurement errors of the item RP 4 and item RE 6. It was hypothesized that the crossloading had been influenced by the same wording of both items RP 4 (concerning physical health) and RE 6 (concerning mental health) ($r_{(RP4-RE6)} = .68$; cf. table 3).

The goodness of fit statistics yield a good to very good model conformity just for the 2nd and 3rd model. Model 1 (without any covariations of errors) fails the goodness of fit tests. Therefore the specification of correlating error terms is relevant for the model fit. Model 3 shows in comparison to model 2 a further improvement of all fit indices.

The figures show the standardized loading for all models, the squared correlations and the latent correlation between the first order factors physical component (PC) and mental component (MC).

A detailed look on model 3 shows that the effect sizes of the paths altogether vary around medium to high levels $\lambda_{x,\xi} > .50$. The latent correlation amounts to $\phi = .83$ and confirms the

Table 2:
Summary of fit statistics of the specified models

fit indices	model 1	model 2	model 3
χ^2; df; Ratio	333.9; 53; 6.3	185.4; 49; 3.8	132.8; 48; 2.8
GFI	.84	.92	.94
AGFI	.77	.88	.90
NFI	.85	.92	.94
TLI	.84	.91	.95
CFI	.87	.94	.96
PGFI	.57	.58	.58
PNFI	.68	.68	.68
SRMR	.06	.05	.04
RMSEA (95 %-CI)	.12 (.11-.14)	.09 (.07-.10)	.07 (.06-.09)

GFI: Goodness of Fit Index; AGFI: Adjusted Goodness of Fit Index; NFI: Normed Fit Index; TLI: Tucker-Lewis Index; CFI: Comparative Fit Index; PGFI: Parsimony Goodness of Fit Index; PNFI: Parsimony Normed Fit Index; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation.

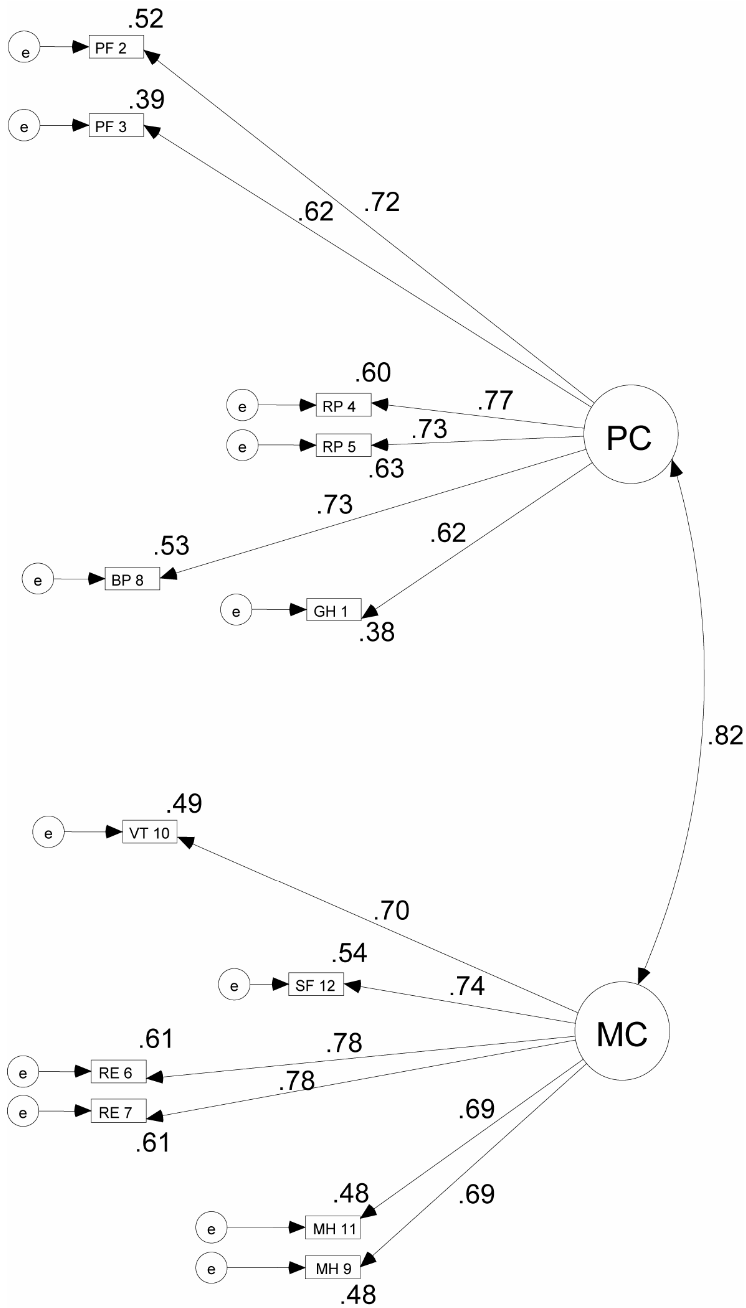

Figure 1:
Standardized solution of the confirmatory factor analysis model 1
(PC = physical component; MC = mental component)

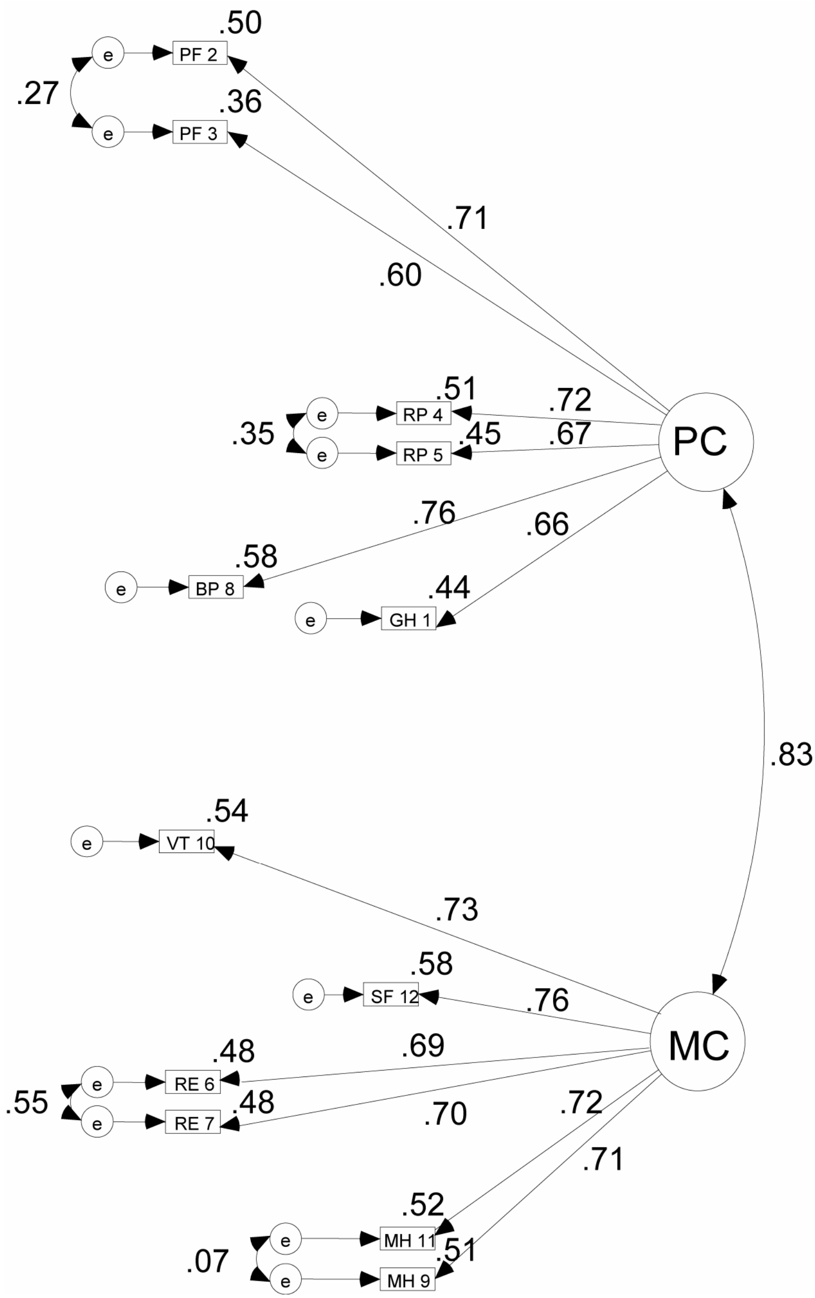

Figure 2:
Standardized solution of the confirmatory factor analysis model 2
(PC = physical component; MC = mental component)

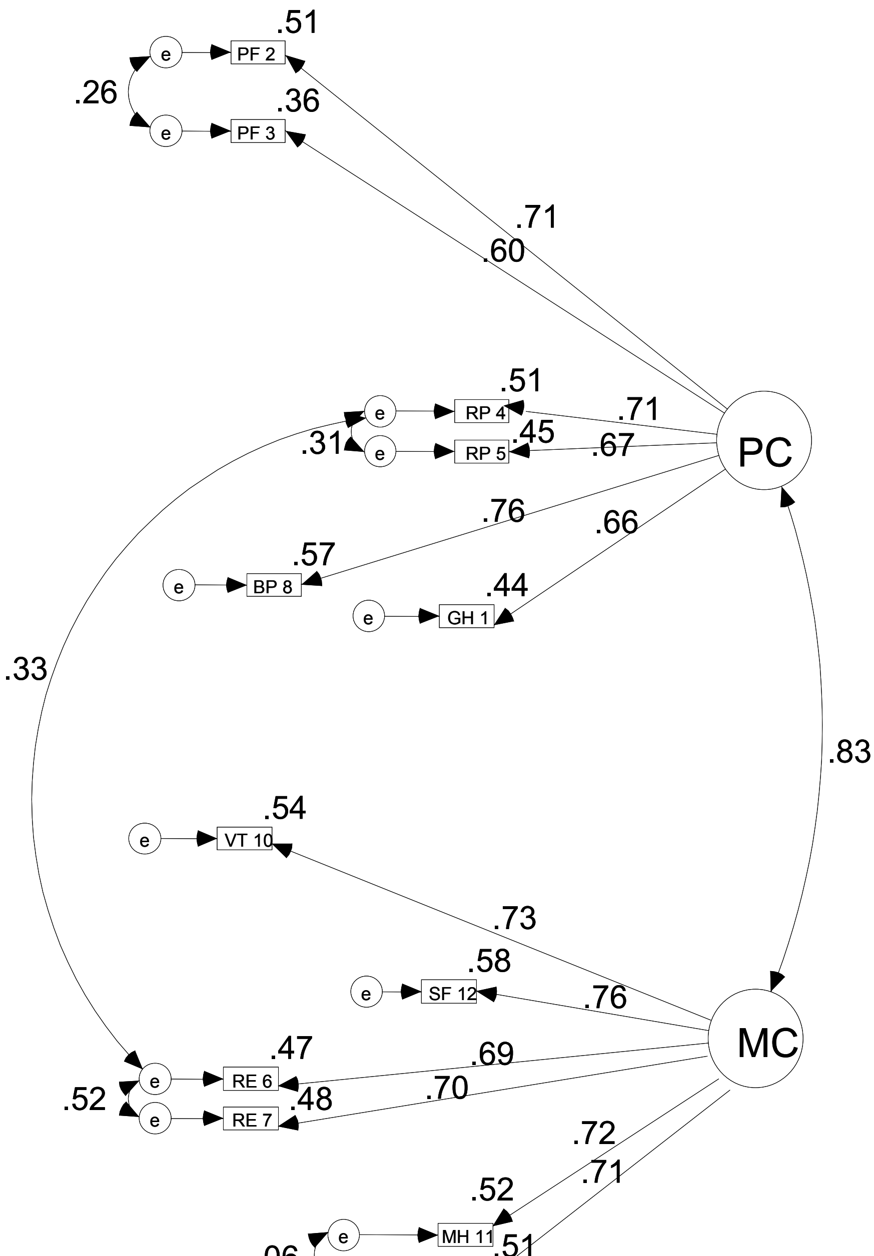

Figure 3:
Standardized solution of the confirmatory factor analysis model 3
(PC = physical component; MC = mental component)

assumed strong association between physical and mental health. The correlation of the error terms within the same subdimension, for example RE 6 and RE 7, are partially high (θ>.50), but for other items, e.g. MHI 10 and MHI 11, they are lower (θ< .10). For both dimensions effect sizes of the paths from the first order factors to the indicator items are all substantial ($\lambda_{x,\xi}$ > .60). The squared correlations as a measure of the explained variance are altogether acceptable.

Age and gender specific analyses

The intercorrelation between the summary scales of SF-12 amounts to $r = .31$ ($p = .000$). The correlation of the PSS ($\alpha = .81$) and age is $r = -.13$ ($p < .02$) and PSS and gender $r = .07$, for MSS ($\alpha = .83$) and age $r = .10$, and MSS and gender $r = .25$ ($p = .000$). The intercorrelations of the individual 12 items and their relationships to age and gender can be found in table 3.

Age does not correlate with the SF-12 items. One exception is item PF 2 "moderate activities" which correlates negatively with age. A positive relationship of gender is seen throughout all six items indicating mental health and some of the physical health items indicating that men might report a better quality of life.

To investigate gender and age specific effects two two-factorial analyses of variance were undertaken with age and gender as independent factors, and the SF-12 summary scales as dependent variables. Therefore age was divided into four groups (lowest through 49 years; 50 to 54 years; 55 to 59 years; 60 through highest). Table 4 shows the results.

Table 3:
Intercorrelations of the items of the SF-12, and with age and gender

	GH 1	PF 2	PF 3	RP 4	RP5	RE 6	RE 7	BP 8	VT 10	MH 9	MH 11	SF 12
PF 2	.45											
PF 3	.40	.56										
RP 4	.45	.52	.44									
RP 5	.36	.53	.42	.66								
RE 6	.38	.36	.37	.68	.50							
RE 7	.37	.38	.35	.56	.49	.77						
BP 8	.54	.56	.44	.50	.52	.42	.40					
VT 10	.49	.42	.39	.43	.32	.49	.48	.46				
MH 9	.38	.36	.29	.40	.35	.45	.49	.36	.61			
MH 11	.41	.38	.27	.44	.35	.50	.48	.38	.53	.57		
SF 12	.42	.45	.36	.51	.45	.50	.53	.46	.50	.55	.59	
PSS §	.66	.76	.67	.67	.69	.41	.36	.84	.49	.32	.33	.48
MSS §	.47	.33	.27	.52	.38	.75	.74	.39	.70	.75	.79	.76
age	-.09	-.11	-.10	-.03	-.06	-.01.	.02	-.07	.02	.10	.08	.02
gender #	-.01	.13	.13	.20	.07	.20	.13	.06	.28	.18	.23	.14

gender: women = 1, men = 2; § PSS = physical health summary scale, MSS = mental health summary scale

Table 4:
Results of two two-way analyses of variance with age and gender as independent factors

	physical summary scale				mental summary scale			
Source	df	*F*	η^2	*p*	df	*F*	η^2	*p*
gender	1	3.8	.01	,05	1	14,1	,04	,00
age	3	3.9	.03	,01	3	3,6	,03	,01
gender x age	3	3.3	.03	,02	3	,64	,01	,59
within group error	335				335			

Concerning the physical summary scale overall (effects of age and the interaction of gender and age are significant: women, and especially older women (group 3 and 4) show less physical quality of life. According to the mental summary scale gender and age main effects are significant: men show higher values than women, and people in group three seem to have the lowest values. However, multiple post hoc comparisons using Scheffe-adjustment do not yield any significant difference between the subgroups.

Discussion

The objective of this study was to examine the psychometric properties of the generic instrument Short Form 12 in patients with diabetes and to test the questionnaire's scale structure using confirmatory structural equation modeling. For this purpose data were pooled from two rehabilitation research projects. Furthermore, age and gender specific correlations for both summary scales of physical and mental health were calculated.

First it must be stressed out, that the method of exploratory factor analyses used here could be criticized. For example, because SF-12 items are measured on different scale levels (e. g. dichotom, ordinal, rating scale) the calculation of one common correlation matrix might be misleading. However, this study does not investigate the best way of dealing with different correlation coefficients, but we tried to reproduce international findings within a diseases-specific sample using a comparable methodological approach. The analyses show that altogether the hypothetical structure with two latent components of the short Form 12 could be supported. The factor loadings of the explorative factor analysis correspond to the range of magnitude which would be expected according to the literature. An exception is the item rolph4 which asks, if the respondent has “accomplished less” because of his physical health. This item loads as well on the mental health component (see table 1). This convergence is even approved by the high intercorrelation (r = .68) of rolph4 and rolem6 (see table 3), but rolem6 asks if the respondent has “accomplished less” because of mental health status. In addition the common variance of both items is evident in the high covariations of their error terms, which leads to an improved model fit via SEM. Therefore the results concur with the results of Wilson et al. (2002), who show that it would be necessary to consider accordant specifications in the SF 12. In respect of content this result means that not alone the mental or physical state determine the kind of answering to these items, but an methodological artifact which seems to lie within the identically wording of the items. Thus it is conjecturable that patients here do not distinguish between physical and mental causes, but respond to the general concept of accomplishment. This assumption that patients do not make a big difference concerning their mental and physical status is supported by the intercorrelations of the latent factors, that are extremely high with $\phi = .83$. This approves the intuitive supposition that mental and physical health is associated very closely in the patients' view. On the manifest level of the summary scales this strong relationship is not evident. This divergence is due to the specific algorithm of the SF-12 summary scale calculation that used orthogonal rotation based on a US norm population sample to calculate the summary scores. So the summary scores are forced to be uncorrelated, which is quite questionable. This method of calculating the summary scales was already criticized within the Short Form 36 literature (Simon et al., 1998; Taft et al., 2001). Johnson and Maddigan (2004) used an alternative scoring function (RAND-12). The RAND-12 employs scaling procedures based on the item response theory and it uses oblique factor rotations to generate mental and physical health summary scores. The RAND-12 has been found to be superior to the SF-12 original algorithm in diabetes type 2 sample, since the SF-12 summary scores have overlooked differences in health status.

The divergence of the high latent dependency of the factors and the low correlation of the SF-12 summary scales raises doubts about the construct validity of the scales as measures of the factors. Construct validity requires that the correlation among latent factors is very similar to the corrected among the instruments for measuring these factors. For practical use the interpretation of the summary scales in patients with diabetes should be performed with caution: mental and physical health cannot be seen as independent, although the summary scales might not be correlated in the empirical sample.

Additional age and gender specific analyses show, that there are low correlations between both summary scales and age. Gender shows a middle correlation with the mental summary scale. Subgroup analyses reveal that women between 50 and 59 years might have the lowest physical quality of life. Men show higher values on the mental summary scale. However, after applying post hoc tests adjusted for multiple comparisons no significant differences remain. Moreover, because Diabetes type I and type II patients were analyzed together in this study, this results should be interpreted with caution.

Additional influences towards the quality of life, for example income or social status (Camacho et al., 2002) need to be further examined in the future. Although the number of missing values is low (about 6 %), it is essential to state the high susceptibility to missing values within the application of the SF 12. Through this an intensive monitoring (consultation with patients) may be necessary. A computer based (German) version of the SF 12 for clinical use (Ryan et al., 2002) would be meaningful, indeed.

In general, the generic SF-12 questionnaire shows factorial validity in patients with diabetes mellitus comparable to international results, whereas the construct validity of the SF-12 has to be criticized. Generic instruments are most useful for cross-illness comparisons of treatments, whereas the benefit of using disease-specific questionnaires lies in the higher sensitivity to specific changes in patients functional capacity or lifestyle issues (Beaser et al, 1996). According to the multidimensional concept of quality of life, that mostly combines physical constitution, mental well-being, social functional, and functional capacity, a generic instrument could measure all these dimensions within different illnesses. Using a disease specific instrument disease-related factors of quality of life could be measured that often have a higher relevance, e. g. vocational status, satisfaction with appearance, or the impact of complications on everyday life. However, if the SF-12 is used in addition to a disease specific instrument, the researcher must be aware that the two latent components are highly dependent even though the summary scores are not or less correlated, and that item response could be influenced more by wording than subjective health from the patients' point of view. Therefore an alternative scoring approach, for example the RAND-12, could be taken into account.

Acknowledgement

This work was funded by the Bundesministerium für Bildung und Forschung, project: Z2.2, No: 01GC0110: "Analyses and Utility of the Short Form 36/Short Form 12 as an Outcome Measurement for Medical Rehabilitation", North German Network for Rehabilitation Research.

References

Amir, M., Lewin-Epstein, N., Becker, G. & Buskila, D. (2002). Psychometric properties of the SF-12 (Hebrew version) in a primary care population in Israel. *Medical Care, 40,* 918-928.

Anderson, R. T., McFarlane, M., Naughton, M. J. & Schumaker, S. A. (1996). Conceptual issues and considerations in cross-cultural validation of generic health-related quality of life instruments. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 605-612). Philadelphia: Lippincott-Raven Publishers.

Beaser, R. S., Garbus, S. B. & Jacobson, A. M. (1996). Diabetes mellitus. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 983-991). Philadelphia: Lippincott-Raven Publishers.

Bott, U., Mühlhauser, I., Overmann, H. & Berger, M. (1998). Validation of a diabetic-specific quality of life scale for patients with type 1 diabetes. *Diabetes Care, 21,* 757-769.

Camacho, F., Anderson, R. T., Bell, R. A., Goff, D. C., Duren-Winfield, V., Doss, D. D. & Balkrishnan, R. (2002). Investigating correlates of health related quality of life in a low-income sample of patients with diabetes. *Quality of Life Research, 11,* 783-796.

DCCT Research Group. (1988). Reliability and validity of a diabetes quality of life measure (DQOL) for the diabetes control and complications trial (DCCT). *Diabetes Care, 11,* 725-732.

Gandhi, S. K., Salmon, J. W., Zhao, S. Z., Lambert, B. L., Gore, P. R. & Conrad, K. (2001). Psychometric Evaluation of the 12-item Short Form Health Survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clinical Therapeutics, 23,* 1080-1098.

Garratt, A. M., Schmidt, L. & Fitzpatrick, R. (2002). Patient-assessed health outcome measures for diabetes: a structured review. *Diabetes Medicine, 19,* 1-11.

Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modelling, 6,* 1-55.

Hu, L., Bentler, P. M. & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351-362.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J. & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18,* 667-683.

Hurst, N. P., Ruta, D. A. & Kind, P. (1998). Comparison of the MOS short form 12 (SF-12) health status questionnaire with the SF-36 in patients with rheumatoid arthritis. *British Journal of Rheumatology, 37,* 862-869.

Jenkinson, C. & Layte, R. (1997). Development and testing of the UK SF-12. *Journal of Health Service Research Policy, 2,* 14-18.

Johnson, J. A. & Coons, S. J. (1998). Comparison of the EQ-5D and SF-12 in an adult US sample. *Quality of Life Research, 7,* 155-166.

Johnson, J. A. & Maddigan, S. L. (2004). Performance of the RAND-12 and SF-12 summary scores in type 2 diabetes. *Quality of Life Research, 13,* 449-456.

Kaplan, R. M. (2003).The significance of quality of life in health care. *Quality of Life Research, 2,* 3-16.

Lim, L .L. Y. & Fisher, J. D. (1999). Use of the 12item Short Form (SF-12) health survey in an Australian heart stroke population. *Quality of Life Research, 8,* 1-8.

Norris, S. L., Engelgau, M. M. & Narayan, K. M. (2001). Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. *Diabetes Care, 24,* 561-587.

Ogasawara, H. (2003). Correlations among maximum likelihood and weighted/unweighted least square estimators in factor analysis. *Behaviormetrika, 30,* 63-86.

Price, P. & Harding, K. (2000). The impact of foot complications on health-related quality of life in patients with diabetes. *Journal of Cutaneous Medicine and Surgery, 4,* 45-50.

Reed, J. P. (1998). Medical outcome study Short Form 36: Testing and cross-validation a second-order factorial structure for health system employees. *Health Service Research, 33,* 1361-1380.

Ryan, J. M., Corry, J. R., Attewell, R. & Smithson, M. J. (2002). A comparison of an electronic version of the SF-36 general health questionnaire to the standard paper version. *Quality of Life Research, 11,* 19-26.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J. & Osher, F. C. (2000). Reliability and validity of the SF-12 Health Survey among people with severe mental illness. *Medical Care, 38,* 1141-1150.

Simon, G. E., Revicki, D. A., Grothaus, L. & von Korff, M. (1998). SF-36 summary scores – Are physical and mental health truly distinct? *Medical Care, 36,* 567-572.

Taft, C., Karlsson, J. & Sullivan, M. (2001). Do SF-36 summary component accurately summarize subscale scores? *Quality of Life Research, 10,* 395-404.

Ware, J. E. (2000). SF-36 Health Survey Update. *Spine, 25,* 3130-3139 .

Ware, J. E., Kosinski, M. & Keller, S. D. (1996). A 12-Item short-form health survey – construction of scales and preliminary tests of reliability and validity. *Medical Care, 34,* 222-233.

Wiesinger, G. F., Pleiner, J., Quittan, M., Fuchsjager-Mayrl, G., Crevenna, R., Nuhr, M. J., Francesconi, C., Seit, H. P., Francesconi, M., Fialka-Moser, V. & Wolzt, M. (2001). Health related quality of life in patients with long-standing insulin dependent (type 1) diabetes mellitus: benefits of regular physical training. *Wiener Klinische Wochenschrift, 113,* 670-675.

Wilson, D., Tucker, G. & Chittleborough, C. (2002). Rethinking und rescoring the SF-12. *Sozial- und Präventivmedizin, 47,* 172-177.

Woodcock, A. J., Julious, S. A., Kinmonth, A. L. & Campbell, M. J. (2001). Problems with the performance of the SF-36 among people with type 2 diabetes in general practice. *Quality of Life Research, 10,* 661-670.