Health-related quality of life instruments and individual diagnosis - a new area of application

Instrumente gesundheitsbezogener Lebensqualität und Individualdiagnostik - ein neuer Anwendungsbereich

Abstract

Objectives: Various health-related quality of life (HRQoL) instruments for children and adolescents have been developed and are applied in clinical and health economic studies with the research focus on the HRQoL of specific groups of responders, comparing these groups respectively. However, if HRQoL assessment aims to contribute on deciding upon a suitable individual treatment and the evaluation thereof, as well as to identify hidden morbidities, an individual diagnostic perspective is required.

To date, the majority of HRQoL questionnaires for children and adolescents are not designed for individual diagnostic assessment and comparison. This paper aims to demonstrate that should a particular measurement tool meet specific psychometric requirements – an individual diagnostic assessment on mental health related problems is possible. We investigated whether or not the KIDSCREEN-27 HRQoL instrument for children and adolescents is able to detect mental health problems in general, provided we apply a principal component analysis (PCA) for summative scaling.

Methods: The cross-sectional survey of the KIDSCREEN project was carried out in 13 European countries (AT, CH, CZ, DE, EL, ES, FR, HU, IE, NL, PL, SE, UK). The Rasch-scaled KIDSCREEN-27 test-data of 22,830 children and adolescents were analysed. To achieve a summative scaling PCA was performed on the correlation between the KIDSCREEN-27 scores. The principal component (PC) values were estimated and used to score the respondents. The reliability and diagnostic quality of this scoring was examined.

Results: The first PC accounted for 58.28% of the entire variance within the KIDSCREEN-27 scores. All KIDSCREEN scales loaded high on the first PC. The reliability of the linear combination of KIDSCREEN-27 scores with the PC-values was $r_{PC}=.94$ and thus above the threshold for individual comparison. This value was higher than the values for the original single KIDSCREEN-27 scores ($r_{single}=.78-.84$). The receiver operating characteristic curve (ROC) was calculated whilst screening for respondents with a mental health problem in general (Strengths and Difficulties Questionnaire). The area under the ROC (AuC) was .80 and statistically significant lower than the AuC issued after a logistic regression analysis employing the 5 original KIDSCREEN-27 scores (AuC=.83). However, according to international conventions, both AuCs denote a "good" discrimination.

Conclusion: Specifically for the KIDSCREEN-27 the results show that the application of a specific scoring algorithm leads to fulfil pre-specified demands of high reliability. The applied scoring approach leads to a good discrimination of the measurement, thus enabling to detect a mental health problem in general using the HRQoL test-data. The presented approach can enhance the usability and range of application of HRQoL measurement.
Zusammenfassung

Zielsetzung: Diverse Instrumente zur Messung der gesundheitsbezogenen Lebensqualität (HRQoL) von Kindern und Jugendlichen wurden entwickelt und werden in klinischen und gesundheitsökonomischen Studien eingesetzt, mit dem Forschungsschwerpunkt auf der Untersuchung der HRQoL von Gruppen, bzw. entsprechenden Gruppenvergleichen. Sofern die Erfassung der HRQoL zur Entscheidung über- und der Bewertung von individueller Behandlung sowie zur Identifizierung nicht entdeckter Morbidität beitragen soll, wird jedoch eine individualdiagnostische Perspektive benötigt. Die Mehrheit der heutzutage verwendeten HRQoL Instrumente für Kinder und Jugendliche sind nicht für Individualdiagnostik ausgelegt. In dieser Arbeit wird demonstriert, dass sofern ein Messverfahren bestimmte psychometrische Kriterien erfüllt, eine brauchbare Individualdiagnostische Erfassung von mentalen Problemen möglich ist. Speziell für den KIDSCREEN-27 HRQoL Fragebogen für Kinder und Jugendliche wurde untersucht ob die Anwendung einer Hauptkomponentenanalyse (PCA) zur summatischen Skalierung, die Identifizierung von mentalen Gesundheitsproblemen ermöglicht.

Methodik: Die als Querschnitterhebung konzipierte Studie des KIDSCREEN Projektes wurde in 13 Europäischen Ländern (AT, CH, CZ, DE, EL, ES, FR, HU, IE, NL, PL, SE, UK) durchgeführt. Die Rasch-skalierten KIDSCREEN-27 Testdaten von 22.830 Kindern und Jugendlichen wurden analysiert. Zwecks summatischer Skalierung wurde eine PCA der Korrelationen zwischen den KIDSCREEN-27 Skalen durchgeführt. Die Hauptkomponentenwerte (PC) wurden berechnet und zur Skalierung der Befragten verwendet. Die Reliabilität und diagnostische Qualität dieses Skalierungsansatzes wurde untersucht.

Ergebnisse: Die erste PC erklärte 58,28% der Gesamtvarianz der KIDSCREEN-27 Skalen. Alle KIDSCREEN-27 Skalen lagen hoch auf der ersten PC. Die Reliabilität der linearen Kombination der KIDSCREEN-27 Skalen zu PC Werten war $r = .94$ und lag damit über dem Grenzwert für individuelle Vergleiche und war höher als die Reliabilität der ursprünglichen einzelnen KIDSCREEN-27 Skalen ($r = .78 - .84$). Die Receiver Operating Characteristic Curve (ROC) wurde berechnet beim Screening nach Befragten mit mentalen Gesundheitsproblemen (Strengths and Difficulties Questionnaire). Die Fläche unter der ROC (AuC) betrug .80 und war statistisch signifikant kleiner als die AuC berechnet nach einer logistischen Regression unter Verwendung der fünf ursprünglichen KIDSCREEN-27 Skalen. Entsprechend internationaler Konventionen können jedoch beide AuCs als ”gute“ Diskrimination bewertet werden.

Fazit: Spezifisch für den KIDSCREEN-27 zeigten die Resultate dass ein spezifischer Skalierungsansatz dazu führt die a-priori spezifizierten Ansprüche einer hohen Reliabilität zu erfüllen. Der angewandte Skalierungsansatz führt zu einer guten Diskrimination der Messung und ermöglicht die Identifizierung von mentalen Gesundheitsproblemen durch HRQoL Testdaten. Der vorgestellte Ansatz kann die Brauchbarkeit und die Anwendungsbreite der Lebensqualitätsmessung erweitern.

Introduction

The assessment of children and adolescents health-related quality of life (HRQoL) in clinical and health economics studies is receiving an increasingly meaningful role as “Outcome” [1], [2]. The term HRQoL denotes, in psychological terminology, a multidimensional construct covering physical, emotional, mental, social and behavi-
HRQoL considers the perceptions of those involved relevant in many aspects, for example the use of health services [8]. Questions about the HRQoL have gained importance in the field of therapy evaluation, especially in treatment of chronic degenerative illnesses that do not lead to a shortening of life, but rather accompany those affected for a lifetime [9]. By concentrating on the needs of those affected we are able to gain additional knowledge about the sense and use of corresponding treatment procedures [10]. Furthermore it reveals subjective differences between therapies.

Various HRQoL instruments for children and adolescents have been developed during the last decade and were applied in clinical and health economic studies with the research focus on the HRQoL of specific groups of responders, comparing these groups respectively. However, if HRQoL assessment aims to contribute to the decision on and evaluation of individual treatment [11] as well as to the identification of hidden morbidities [12], an individual diagnostic perspective is required. Yet the question arises whether or not the current HRQoL questionnaires for children and adolescents can be used for individual diagnostic assessment as well.

To date, the majority of HRQoL questionnaires for children and adolescents are not designed for individual diagnostic assessment. For example, the interpretation of the scores via reference data is focussed on group comparison by e.g. providing cut-off values which inform that there is a noticeable deviation from the reference population when a specific group of responders display only moderately lower HRQoL on average than the overall reference population. However, for individual diagnosis lower thresholds are required. An exception is the generic KIDSCREEN HRQoL Instrument [13] which was developed simultaneously in several different countries and tested in a large representative sample of children and adolescents. Issued from the field testing, reference data and cut-off values for the interpretation of individual KIDSCREEN scores are provided [14]. However, individual diagnostic assessment demands the test-scores of an instrument to be highly precise. Such high reliability should ensure a small measurement error confidence interval, necessary to identify statistically significant differences in the HRQoL trait parameter values when comparing individuals. Reliability of .90 or even .95 has been proposed as opposed to lower acceptable reliability of at least .70 for group comparison [15]. Most of the HRQoL scales for children and adolescents achieve a satisfying reliability of .70 and more, but do not meet such strict requirements [16]. This is demonstrated profoundly in profile instruments where the items are clustered into many different scales.

This paper aims to demonstrate that should a particular measurement tool meet several psychometric requirements; a diagnostic assessment on mental health related problems is possible. For the KIDSCREEN HRQoL instrument [13] specifically we investigated, if the application of a principal component analysis (PCA) [17] for summative scaling [18] meets the psychometric demands for individual diagnostic assessment of mental health related problems. The diagnostic quality of the resulting summative scaling will be examined when screening for children and adolescents with mental health problems. The discriminating ability will be compared with the discrimination achieved when using the information of the complete KIDSCREEN-27 measurement profile.

Methods

Subjects and study design

The KIDSCREEN project was funded by the European Commission. It aimed to develop the first instrument using a simultaneous international construction approach which includes the exploration of the relevant dimensions of HRQoL in the questionnaire, considering different age groups and different countries. Furthermore the initial scale content should be based on content generated by focus groups in different countries [19]. Modern techniques for the development and cross-cultural validation of HRQoL, such as structural equation modelling and the item response theory were applied [20]. The result of this project was the generic KIDSCREEN HRQoL questionnaire [13].

The cross-sectional KIDSCREEN survey was carried out to screen and promote the HRQoL in children and adolescents in European countries from a public health perspective. National representative surveys on health and quality of life were conducted in the 13 participating European countries. Three different approaches to sample selection were used. In six countries (AT, DE, CH, ES, FR, NL), address sampling was conducted via computer-assisted telephone interviews. Questionnaires were sent by post to families who had agreed by phone to participate, and these were filled in at home and sent back to the national centres in a prepaid envelope. In five countries (EL, HU, IE, PL, SE), the samples were obtained in schools that were representative for the country as a whole in terms of type (private vs. public, rural vs. urban). The pupils filled in the questionnaires during class-time and took a copy of the parents’ questionnaire back home with them (except PL, which followed post administration). Parents were asked to fill in the questionnaires and to send them back to each national centre using a prepaid envelope. UK combined both telephone and school administration and CZ used a multistage random sampling of communities and households. Design, sampling and administration methods by country, as well as assessment of the external validity are described in detail elsewhere [21].
Instruments and variables

The generic KIDSCREEN-27 was developed from the extended KIDSCREEN-52 HRQoL questionnaire for children and adolescents [13]. The instrument consists of 27 items assessing five HRQoL dimensions: Physical Well-Being explores the level of the children/adolescent’s physical activity, energy and fitness. Psychological Well-Being examines the psychological well-being of the child/adolescent, including positive emotions, satisfaction with life and feeling emotionally balanced. Parent Relations & Autonomy examines relationships with parents, the atmosphere at home and feelings of having enough age-appropriate freedom to choose (things for yourself in the relationship, good balance between parents) as well as feeling satisfied with the own financial resources. Social Support & Peers examines the nature of the respondents’ relationships with other children/adolescents. Finally School Environment explores the child’s/adolescent’s perceptions of his/her cognitive capacity, learning and concentration, and their feelings about school.

The items of the KIDSCREEN-27 HRQoL questionnaire assess either the frequency of behaviour/feelings or, in fewer cases, the intensity of an attitude. Both possible item formats use five-point answer categories, with an one week recall period. Negatively worded items were recoded and the item-scores summed up. Because the KIDSCREEN items fulfil the assumptions of the probabilistic partial credit model [20] Rasch-scores are computed for each dimension and are transformed into T-Values with a mean of 50 and a standard deviation of 10; higher scores indicate higher HRQoL and well-being. The international manual of the KIDSCREEN instrument (accessible under http://www.kidscreen.org) provides international and national reference data stratified for age and gender. In addition to the reference data for group comparison there are detailed reference values for individual diagnosis as well.

Mental health problems were assessed with the Strength and Difficulties Questionnaire (SDQ) a brief behavioural screening questionnaire [22]. The SDQ explores positive and negative attributes in 25 items regarding emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. Additionally, information on the impact of a potential mental condition is gathered. The SDQ-self and parents report version was included as a screening instrument of mental health in all countries except Ireland. The SDQ was scored using the “predictive algorithm” which aims to predict any psychiatric diagnosis in the categories “unlikely”, “possible” and “probable” using the self-report and parent report as well as their information on the impact [23], [24]. However, we have to bear in mind that the SDQ can only determine whether a subject has a mental health problem or not. Although the SDQ represents a dimensional assessment tool, only the three-categorical outcome of the predictive algorithm was used, because its combination of the different data-sources could be viewed as being more valid than the single SDQ-scores [25]. For the analyses the variable was dichotomized into the collapsed categories “unlikely and possible” (value = 0) and “probable” (value = 1). Fewer parents than children and adolescents participated the KIDSCREEN survey. Thus only 18,625 children and adolescents could be categorized according to the SDQ predictive algorithm.

Sociodemographic information on e.g. age, gender and nationality were collected via self report from the children and adolescents.

Statistical analysis

The reliability of the KIDSCREEN-27 scores was calculated with Cronbach’s alpha [26]. To achieve a summative scaling of the KIDSCREEN-27 scores principal component analysis (PCA) was performed on the matrix of Pearson-correlation between the KIDSCREEN-27 scores. The principal component (PC) values were estimated and used to score the respondents (KIDSCREEN-27 PC-score). This approach has been used in the HRQoL testing community for summative scaling of HRQoL measures before [27]. However, only the first (general) PC was extracted and the PC scores calculated - an approach also described by Chen [18] for unidimensionalized scaling of multidimensional (meta-) constructs.

The percentage of total variance was explained and the factor loadings were computed. The eigenvalue plot was examined to justify the general factor solution. The distribution of the PC-scores was examined. To compute the reliability of the PC-scores a formula for the reliability of a linear combination [15] was used: The reliabilities of the KIDSCREEN-27 scores were multiplied with their corresponding squared PC-scoring coefficients and summed up. This term is subtracted from the sum of the squared PC-scoring coefficients and the result divided by the variance of the PC-values. Subtracting the last value from 1 equals the reliability of the linear combination used for calculating the PC-scores.

The ability of the KIDSCREEN-27 PC-score to detect a mental health problem in general was evaluated by fitting a binary logistic regression model [28] to predict the dichotomized SDQ using the KIDSCREEN-27 PC-score as predictor. Regression coefficients were estimated stratified for age-group (<12 years and >11 years), gender and country. The Nagelkerke pseudo-R² value, the regression coefficients and the odds ratios were computed. The area under the receiver operating characteristic curve (AuC) [29] was calculated when screening for respondents with a mental health problem. For every possible cut-off value the sensitivity and the specificity (distracted from 1) is calculated and plotted in 2-axis coordinates. The area under the plotted line denotes the discrimination. The discrimination indicated by the AuC was evaluated according to the rules of Hosmer & Lemeshow [28]. Values between .70 and .79 were considered as evidence for acceptable discrimination, values between .80 and .89 as good discrimination.

Condensing the KIDSCREEN-27 test data into one single general value of HRQoL (KIDSCREEN-27 PC-score) might
lead to loose crucial information about the respondents HRQoL. In order to track potential loss of such psychometric information, the (AuC) achieved via KIDSCREEN-27 PC-score was compared with the AuC issued after fitting a logistic regression model to predict the dichotomized SDQ using the five original KIDSCREEN-27 scores as predictors. The regression coefficients were again estimated stratified for country, age and gender. The two AuCs were compared and tested for statistically significant deviation. Fitting a multiple logistic regression model encompasses a linear combination of the predictors in the exponent of the logistic regression formula. Thus the reliability of this linear combination was calculated too using a similar formula like the one described above [15]. The statistical analyses were carried out using the software SPSS 13.0 and Stata 9.0.

Results

Sample description

The final sample included 22,830 children and adolescents. Response rates varied across countries, from 45.3% to 100%. Table 1 shows the socio-demographic characteristics of the final sample, overall and by country. The mean age for the overall child sample (8-11 years) was 9.6 years, and for the adolescent sample (12-18 years) 14.3 years. There were slightly more females than males in both samples. In terms of age and gender, the child and adolescent samples were broadly similar across all participating countries. The most notable differences between countries occurred in socio-economic status, assessed with the family affluence scale (FAS [30]) with, for example, 49.5% of the Czech Republic child sample reporting low FAS compared to only 7.5% in the French sample. A similar pattern was seen in the adolescent sample.

Psychometric properties of the KIDSCREEN-27 profile scores

The internal consistency reliability of the KIDSCREEN-27 HRQoL questionnaire dimension scores for the overall sample as well as the range of their internal consistencies across countries was calculated. In the overall sample Cronbach’s alpha ranged from .78 to .84. None of the scales displayed Cronbach’s alpha below .70 in any of the country subgroups. Due to the scoring algorithm the scale scores were distributed with means around 50 and standard deviations (SDs) around 10. The percentage of respondents omitting at least one item in a scale ranged from 1.57% (Social Support & Peers) to 3.80% (Parents & Autonomy).

Principal component analysis

Aimed to achieve a summative scaling of the KIDSCREEN-27 instrument, PCA was performed on the matrix of Pearson-correlation between the KIDSCREEN-27 scores. Beforehand this correlation matrix was examined. The inter-scale correlation ranged from r=.37 to r=.59. The first PC was extracted. This PC accounts for 58.28% of the whole variance in the KIDSCREEN-27 scores. The examination of the scree-plot of eigenvalues (Figure 1) indicated a one (general-) PC solution: A large eigenvalue (2.91) for the first PC was followed by a sharp decrease to the eigenvalues of the succeeding PCs (.66 to .35). Furthermore, the eigenvalue of the second PC was below criterions specified in the literature [17]. The PC-loading coefficients of the KIDSCREEN-27 profile scores ranged from .68 (Social Support & Peers) to .86 (Psychological Well-being) (Table 2).

Convergence with mental health problems/diagnostic quality

The ability of the KIDSCREEN-27 PC-score to detect a mental health problem in general was evaluated by fitting a binary logistic regression model to predict the dichotomized SDQ using the KIDSCREEN-27 PC-score as predictor. Regression coefficients were estimated stratified for age-group (<12 years and >11 years), gender and country. The Nagelkerke pseudo-R² was .16. The odds ratio of KIDSCREEN-27 PC-score and probable mental health problems was .31 indicating that the higher the HRQoL – the less likely (chance) the respondents were to display a mental health problem (Table 3). The corresponding 95% confidence interval (.29-.34) did not include the value of 1.

The area under the receiver operating characteristic curve (AuC) [29] was calculated when screening for respondents with a mental health problem in general (SDQ = probable) using the KIDSCREEN-27 PC-score. For every possible cut of value the sensitivity and the specificity (distracted from 1) was calculated and plotted in a 2-axis coordination (Figure 3, blue line). The area under the plotted line (AuC) was .80, according to the rules of [27] denoting a good discrimination. For a sensitivity of .80(.75) the specificity would be .62(.70). For a specificity of .80(.75) the sensitivity would be .65(.70).

The discrimination achieved with the original KIDSCREEN-27 profile scores was examined by first fitting a logistic regression model to predict the dichotomized SDQ using the KIDSCREEN-27 profile score as predictors. Regression coefficients were estimated stratified for age-groups (<12
Table 1: Response rate and socio-demographic characteristics of the KIDSCREEN sample

| Gender | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
|--------|------|------|------|------|------|------|------|------|------|
| Children | 6985 | 5185 | 5185 | 5185 | 5185 | 5185 | 5185 | 5185 | 5185 |
| N | 2230 | 1475 | 1475 | 1475 | 1475 | 1475 | 1475 | 1475 | 1475 |
| Total Asia, Australia, Swiss, Czech, Germany, Spain, France, Greece, Hungary, Ireland, Lithuania, Poland, Sweden, UK, United Kingdom | 633 | 423 | 423 | 423 | 423 | 423 | 423 | 423 | 423 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
| Female | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Male | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 | 64.7 |
| Age range | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 | 8-11 |
| Mean age yrs (SD) | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 | 14.7 |
| Ranges | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 | 12-18 |
Table 2: Principal component analysis of KIDSCREEN-27 profile scores

| Variables              | Loadings | Coefficients for PC scores | Communality |
|------------------------|----------|----------------------------|-------------|
| Physical Well-Being    | 0.72     | 0.25                       | 0.52        |
| Psychological Well-Being| 0.86     | 0.29                       | 0.73        |
| Parents & Autonomy     | 0.79     | 0.27                       | 0.62        |
| Social Support & Peers | 0.68     | 0.23                       | 0.46        |
| School Environment     | 0.76     | 0.26                       | 0.58        |

N=20596; eigenvalue = 2.91; variance explained = 58.28%

Figure 1: Principal component analysis of KIDSCREEN-27 scores: Scree plot of eigenvalues

Figure 2: Histogram (20 bins) of the distribution of KIDSCREEN-27 PC-scores (mean=0, SD=1)
### Table 3: Logistic regression of mental health problems (SDQ) on KIDSCREEN-27 PC-score

| Independent variable | B   | SE(B) | p     | EXP(B) | KI inf. | KI sup. |
|----------------------|-----|-------|-------|--------|---------|---------|
| KIDSCREEN-27 PC-score| -1.16 | 0.04  | <.001 | 0.31   | 0.29    | 0.34    |
| Constant             | -3.16 | 0.04  | <.001 |        |         |         |

N=18625; Nagelkerke pseudo R² = 0.16

### Figure 3: Analysis of the area under the receiver operating curve: AUC=.80 for PCA-approach (blue line); AUC=.83 for logistic regression approach (red line)

### Table 4: Logistic regression of mental health problems (SDQ) on KIDSCREEN-27 profile scores

| Independent variable     | B      | SE(B) | p     | EXP(B) | KI inf. | KI sup. |
|--------------------------|--------|-------|-------|--------|---------|---------|
| Physical Well-Being      | -0.02  | 0.04  | .618  | 0.98   | 0.91    | 1.06    |
| Psychological Well-Being | -0.60  | 0.05  | <.001 | 0.55   | 0.50    | 0.61    |
| Parents & Autonomy      | -0.26  | 0.05  | <.001 | 0.77   | 0.70    | 0.84    |
| Social Support & Peers   | -0.06  | 0.04  | .103  | 0.94   | 0.88    | 1.01    |
| School Environment      | -0.57  | 0.04  | <.001 | 0.57   | 0.52    | 0.62    |
| Constant                 | -3.23  | 0.04  | <.001 |        |         |         |

N=18625; Nagelkerke pseudo R² = 0.22

### Table 5: Comparing the AuC issued from the KIDSCREEN-27 PC-score with the AuC issued after logistic regression of the original KIDSCREEN-27 profile scores: Screening for mental health problems

| Test-scores               | N    | AuC | SE  | KI inf. | KI sup. |
|---------------------------|------|-----|-----|---------|---------|
| KIDSCREEN-27 PC-scores   | 18625| 0.80| 0.01| 0.79    | 0.82    |
| Logreg predicted probability | 18625| 0.83| 0.01| 0.82    | 0.84    |

years and >11 years), gender and country. The Nagelkerke pseudo-R² was .22. The odds ratio of KIDSCREEN-27 profile scores and probable mental health problems were statistically significant smaller than 1 for the scales Psychological Well-Being (.55; 95CI .50-.61), Parents & Autonomy (.77; 95CI .70-.84) and School Environment (.57; 95CI .52-.62): A higher HRQoL is associated with a lower risk (chance) of mental health problems (Table 4).

The predicted probability of mental health problems was calculated and saved as test-scores. The linear combination of the KIDSCREEN-27 profile scores in the exponent of the logistic formula was examined and the reliability of this linear combination was calculated using a formula described in [14]. Using the reliabilities of the KIDSCREEN-27 profile scores and the logistic regression coefficients a reliability of .92 was calculated. Using the predicted probability of mental health problems as test-data, the AuC was calculated when screening for
respondents with a mental health problem (SDQ = probable). The AuC was .83 (Figure 3, red line) and statistically significant larger than the AuC achieved with the KIDSCREEN-27 PC-score (Table 5). However according to the rules of [28] both values denote a good discrimination.

Discussion

The majority of the HRQoL instruments currently available for children and adolescents are neither designed for individual diagnostic assessment nor do they meet the strict psychometric criterions required for individual comparison. Even instruments with overall good psychometric properties most often do not achieve the necessary measurement precision. This paper aimed on investigating if the application of a PCA approach for summative scaling on the children and adolescents HRQoL instrument KIDSSCREEN-27 enables an individual diagnostic assessment on mental health related problems. The application of a conventional PCA demands continuous and interval-scaled variables which ideally should be multivariate normal distributed. From a methodological point of view the Rasch-scores of the KIDSCREEN-27 could be considered to represent continuous and interval-scaled traits. Univariate examinations (not presented here) showed approximately normal distribution of the scores. Thus the KIDSCREEN-27 met these crucial pre-requisites for the application of the PCA. The results of the PCA show that the correlation between the KIDSCREEN scales could be reasonably explained by a general (PC)-factor. The investigation of the eigenvalue plot revealed no hint to a sizeable secondary factor. All KIDSCREEN-27 scales displayed high loadings on that (PC)-factor which was then interpreted as a general HRQoL factor. These results differ from similar PCA analyses of the adults SF-36 HRQoL measure [27], were a solution with two (secondary-) components (mental- and physical) emerged. However, factor analytic studies on the adults WHOQOL HRQoL measure has lead to a solution with only one secondary factor [31]. Results show that specifically for the KIDSCREEN-27 HRQoL instrument the applied PCA approach leads to a high measurement precision sufficient for individual diagnostic assessment on mental health related problems and individual comparison of test-scores. The distribution of the PC-scores resembled the normal distribution – thus enabling a good discrimination between respondents with different HRQoL across the whole trait to be measured [15]. Using the PC-scores issued from the PCA as test-scores, the question arose about the conceptual justification of such a presentation of HRQoL test-values as a single index. The sizeable intercorrelation might be viewed as a statistical justification – when considering the inter-scale correlation as caused by a latent factor of global HRQoL. In spite of the broadly consensual understanding of HRQoL as a multidimensional construct [3], disagreement exists over, on the one hand, how many and with which dimensions those affected should judge their HRQoL and, on the other, over what meaning the assumed multidimensionality has for the measurement: Because of its multidimensionality, does the HRQoL necessarily have to be inferred over several measurement results? Or can the HRQoL also be fixed through a global value [4]? Bullinger [32] states a consensus already in 1989 that HRQoL can be operationalized through single components and also be globally determinable. Rogerson [33] perceives a broad acceptance in the research community of the presentation of HRQoL measures as an aggregate index as well as a profile. Exact knowledge regarding the dimensionality of an issue is principally of great importance for its adequate assessment. The undervaluing of the number of measured dimensions can either lead to the negligence of meaningful information or to the combination of factors that can then no longer be interpreted [34]. Thus it is important to test if such a global scoring misses important psychometric information.

The investigations on diagnostic quality when trying to detect a mental health problem in general showed that with regards to the AuC, the summative scoring of the KIDSCREEN-27 achieved a similar “good” discrimination as the usage of the whole profile. However the examination of the Nagelkerke pseudo-R² revealed a sizeable higher magnitude of association between test-scores and mental health problems for the latter. These differences hint at an important question: For which specific research aim can a global HRQoL index be used and for which not. This has to be considered theoretically but also must be empirically tested. Thus although the presented results show no sizeable loss of psychometric information when screening for mental health problems, it might be different for other chronic health conditions, and for other research questions respectively. The usefulness of the presented PCA approach has to be explicitly tested for all desired areas of application.

Limitations of the study

Several limitations had to be considered with regards to the presented results. A major limitation concerns the usage of the SDQ for classifying the respondents as having mental health problems or not. We have to bear in mind that the SDQ represents a screening instrument itself too that offers a dimensional assessment of mental health problems, although there are indices that the SDQ is sensitive enough to detect some psychiatric disorders in childhood [23]. Further studies should try to use information on clinical diagnosis to evaluate the discrimination. Further studies should also focus on examining the diagnostic quality whilst screening for mental health problems using a longitudinal design. Another important limitation was rooted in the actual statistical analysis strategy. Both the PCA aimed to develop a summative scoring for the KIDSCREEN-27 profile score as well as the calculation of the diagnostic quality of that scoring were conducted on the same
sample. From a scientific point of view however, the development and testing of a particular scoring should be done using different samples. However as the development of the scoring is based on the internal structure of the KIDSCREEN-27 profiles, whereas the testing of the diagnostic quality focused on the association with an external criterion, the results shown must not be considered sample dependant.

Lastly it has to be considered that the results were issued from a sample of female and male children and adolescents with a large age-range and from a diversity of cultures, differing with regards to various cultural and socioeconomic aspects as well as different sampling designs. While the examination of the diagnostic quality takes into account the heterogeneity of the sample, for the PCA it was decided to solely analyse the whole sample, because the KIDSCREEN scores proved to enable a comparable assessment of HRQoL across gender, age and country [13], displaying no differential item functioning according to the approach described by [35].

A last limitation concerns the applicability of the introduced scoring approach. The reported results solely represent a methodological investigation about the potential of such a scoring. The proposed scaling approach can in no case be viewed as an official scoring for the KIDSCREEN-27. The development of a summative scoring algorithm for the KIDSCREEN is a further task of the KIDSCREEN group and will include the comparison of different approaches, including the one presented.

Conclusions

The usage of HRQoL test data for individual diagnostic assessment and comparison is connected to strict psychometric criteria which are usually not fulfilled by the current available instruments for children and adolescents. However it could be shown that given an instrument fulfill the requirements necessary for the application of the presented PCA analysis (like the KIDSCREEN-27) - the application of the presented scoring approach enables to fulfill the pre-specified demands of a measurement with a high reliability necessary for individual comparison. The reported scoring leads to a good discrimination of the measurement and enable an individual diagnostic assessment on mental health related problems. The presented approach can enhance the usability and range of application of HRQoL measurements.

Notes

Conflicts of interest: none declared

References

1. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: A critical review of published cost-utility studies in child health. Pediatrics. 2005;115 (5):600-14.
2. Koot HM, Wallander M, editors. Quality of Life in Children and Adolescents: Concepts, methods and findings. Rotterdam: Brunner-Routledge; 2001.
3. Bullinger M. Assessing health related quality of life in medicine. An overview over concepts, methods and applications in international research. Restor Neurol Neurosci. 2002;20(3-4):93-101.
4. Schumacher J, Klaiberg A, Brähler E. Diagnostik von Lebensqualität und Wohlbefinden - Eine Einführung. In: Schumacher J, Klaiberg A, Brähler E, editors. Diagnostische Verfahren zu Lebensqualität und Wohlbefinden. Göttingen: Hogrefe; 2003. p. 9-24.
5. Taillfer M, Dupin G, Roberge M, Lemay S. Health-related quality of life models: Systematic review of the literature. Soc Indic Res. 2003;64(2):293-323.
6. WHOQOL Group. The World Health Organization Quality of life assessment (WHOQOL): Position paper from the World Health Organization. Soc Sci Med. 1995;41(10):1403-9.
7. Wasem J, Hessel F. Gesundheitsbezogene Lebensqualität und Gesundheitsökonomie. In: Ravens-Sieberer U, Cleza A, editors. Lebensqualität und Gesundheitsökonomie in der Medizin. Landsberg: Ecomed; 2000. p. 319-35.
8. Seid M, Varni JW, Segall D, Kurtin PS. Health-related quality of life as a predictor of pediatric healthcare costs: A two-year prospective cohort analysis. Health Qual Life Outcomes. 2004;2:48.
9. Radoschewski M, Gesundheitsbezogene Lebensqualität - Konzepte und Maße. Entwicklungen und Stand im Überblick. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2000;43(3):165-89.
10. Patrick DL, Richardson M, Starks HE, Rose MA, Kinne S. Rethinking prevention for people with disabilities, Part II: a framework for designing interventions. Am J Health Promot. 1997;11(4):261-3.
11. Sullivan M. The new subjective medicine: taking the patient’s point of view on health care and health. Soc Sci Med. 2003;56(7):1595-604.
12. Varni JW, Burwinkle TN, Lane MM. Health-related quality of life measurement in pediatric clinical practice: An appraisal and precept for future research and application. Health Qual Life Outcomes. 2005;3(1):34.
13. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W, Auquier P, Abel T, Czimbalmos A, Tountas Y, Hagquist C, Kliroé J & the European KIDSCREEN Group. The KIDSCREEN-52 Quality of life measure for children and adolescents: Development and first results from a European survey. Expert Rev Pharmacoecon Outcomes Res. 2005;5(3):353-64.
14. The KIDSCREEN Group Europe. The KIDSCREEN questionnaires. Quality of life questionnaires for children and adolescents - handbook. Lengerich: Papst Science Publisher; 2006.
15. Nunnally J, Bernstein U. Psychometric theory. 3rd edition. New York: Mc-Graw-Hill; 1994.
16. Rajmil L, Herdan M, De Sanmamed MJF, Detmar S, Bruil J, Ravens-Sieberer U, Bullinger M, Simeoni MC, Auquier P & the European KIDSCREEN Group. Generic Health-related Quality of Life Instruments in Children and Adolescents: A Qualitative Analysis of Content. J Adol Health. 2004;34(1):37-45.
17. Joliffe IT. Principal Component Analysis. 2nd edition. New York: Springer; 2002.
18. Chen S, Sullivan NY, Lu YE, Lubben JE. Unidimensionalization in applied science and social science: A multiple and optimal approach to a key issue in scaling. In: Blasius J, Hox J, de Leeuw E, Schmidt P, eds. Social Science Methodology in the New Millennium (CD-ROM). Leverkusen, Germany: Leske & Budrich; 2001.

19. Ravens Sieberer U, Gosch A, Abel T, Detmar S, Bruijl J, Ravens-Sieberer U, Bullinger M, Simeoni MC, Auquier P & the European Kidscreen Group. Quality of Life in children and adolescents - a European public health perspective. Soz Praventivmed. 2001;46(5):297-302.

20. Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah: Lawrence Erlbaum Associates; 2000.

21. Rajmil L, Berra S, Erhart M, Tebe C, Blæsgger C, Duer W, von Rueden U, Alonso J, Ravens-Sieberer U & the European Kidscreen Group. Methods and representativeness of European surveys on children and adolescents: the KIDSCREEN study. (submitted to BMC Public Health).

22. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38:581-6.

23. Goodman R, Ford P, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Br J Psychiatry. 2000;177:534-9.

24. Goodman R, Renfrew D, Mullick M. Predicting type of disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. Eur Child Adolesc Psychiatry. 2000;9:129-34.

25. Goodman R. The extenden version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry. 1999;40(5):791-9.

26. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;6:297-34.

27. Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of Methods for the Scoring and Statistical Analysis of SF-36 Health Profile and Summary Measures: Summary of Results from the Medical Outcome Study, Med Care. 1995;33(4S):S264-79.

28. Hosmer DW, Lemeshow S. Applied logistic regression (2nd edition). New York: John Wiley & Sons; 2000.

29. Hanley JA, McNeil BJ. The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. Radiology. 1982;143:29-36.

30. Currie CE, Elton RA, Todd J, Platt S. Indicators of socioeconomic status for adolescents: the WHO Health Behaviour in School-aged Children survey. Health Educ Res. 1997;12:385-97.

31. WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQoL): Development and general psychometric properties. Soc Sci Med. 1998;46:1569-85.

32. Bullinger M. Forschungsinstrumente zur Erfassung der Lebensqualität bei Krebs - ein Überblick. In: Verbes B, Hasenbring M, editors. Psychosoziale Aspekte der Krebsforschung (Psychosoziale Onkologie) Jahrbuch der Medizinischen Psychologie. 3rd ed. Berlin: Springer; 1989 p. 45-57.

33. Rogerson RJ. Environmental and health-related quality of life: Conceptual and methodological similarities. Soc Sci Med. 1995;41(10):1373-82.

34. Coste J, Bouée S, Ecosse E, Leplège A, Pouchot J. Methodological issues in determining the dimensionality of composite health measures using principal component analysis: Case illustration and suggestions for practice. Qual Life Res. 2005;14(3):641-54.

35. Zumbo BD. A handbook on theory and methods of differential item functioning (DIF). Ottawa: National Defense Headquarters; 1999.

Corresponding author: PD Dr. Ulrike Ravens-Sieberer Research Unit Psychosocial Health, Robert Koch-Institute, Seestr. 10, 13353 Berlin, Germany, Tel.: +49-30-4547-3436, Fax: +49-30-4547-3531 Ravens-SiebererU@rki.de

Please cite as Erhart M, Ravens-Sieberer U. Health-related quality of life instruments and individual diagnosis - a new area of application. GMS Psycho-Social-Medicine. 2006;3:Doc12.

This article is freely available from http://www.egms.de/en/journals/psm/2006-3/psm000032.shtml

Copyright ©2006 Erhart et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en). You are free: to Share — to copy, distribute and transmit the work, provided the original author and source are credited.