

CLINICAL STUDY

Development and psychometric properties of a disease-specific quality of life questionnaire for adult patients with growth hormone deficiency

P Herschbach, G Henrich, C J Strasburger¹, H Feldmeier¹, F Marín², A M Attanasio³ and W F Blum⁴

Institut und Poliklinik für Psychosomatische Medizin, Psychotherapie und Medizinische Psychologie der Technischen Universität München, München, Germany, ¹Medizinische Klinik Innenstadt, LMU, München, Germany, ²Lilly Spain, Madrid, Spain, ³Eli Lilly and Company, Florence, Italy and ⁴Eli Lilly and Company, Bad Homburg, Germany

(Correspondence should be addressed to P Herschbach, Institut für Psychosomatische Medizin der Technischen Universität München, Langerstr. 3, 81675 München, Germany; Email: p.herschbach@lrz.tum.de)

Abstract

Background: Adults with growth hormone (GH) deficiency (GHD) may experience physical and psychological disturbances, which can affect their quality of life (QOL).

Objectives: To develop and validate a disease-specific module from the previously published QOL measure Questions on Life Satisfaction Modules (QLS^M); the QLS^M-H that specifically addressed the needs of patients with hypopituitarism. A second aim was for the questionnaire to be applicable across different cultural backgrounds in order to evaluate the efficacy of therapy in large, international clinical trials, thus providing additional clinical endpoints for these studies.

Design: A preliminary German language version of the QLS^M-H was developed from 26 semi-structured interviews of adults with GHD. The questionnaire was then independently translated into five other languages and applied in open, non-controlled, multicentre, longitudinal studies to patient ($n = 717$) and normative populations ($n = 2700$).

Methods: A revised, nine-item version of the questionnaire was developed, based on previously defined criteria, and was evaluated for reliability and validity. Sensitivity to detect changes after GH replacement was also assessed.

Results: The 16 items of the preliminary questionnaire were reduced to nine items on the basis of the correlation of items/factors from initial patient interviews. Psychometric analysis revealed the reliability of the nine-item scale. The Cronbach's alpha scores ranged from 0.81 to 0.89 and the test-retest correlations ranged from 0.76 to 0.88, all of which indicate reliability over time. Mean scores increased significantly during GH replacement therapy, with observed changes greater than those seen with the non-specific modules of the QLS^M, indicating the sensitivity of the scale.

Conclusions: The QLS^M-H questionnaire is concise, easy to complete, and can be effectively applied across different cultural backgrounds. Psychometric evaluation of the questionnaire reveals that it is a valid, reliable and sensitive tool useful for assessing impaired life satisfaction in adult patients with GHD and also for monitoring the efficacy of GH therapy.

European Journal of Endocrinology 145 255–265

Introduction

Growth hormone (GH) deficiency (GHD) in adults is associated with a variety of metabolic disturbances that manifest as significant alterations in serum lipid concentrations, bone and glucose metabolism (1), and body composition (2). Furthermore, adults with GHD experience disturbances in their psychological well-being, including factors such as dissatisfaction with body image, low energy levels, emotional lability, mental fatigue and impaired psychosocial status (3, 4). These factors can have a significant impact on a patient's life satisfaction or quality of life (QOL).

Health-related QOL relates to how an individual feels, functions and responds in daily life. The potential to measure a patient's QOL has become increasingly important, not only as a method for predicting the need for treatment, but also to provide additional endpoints in clinical trials designed to assess the efficacy of such therapy. Defining the psychological component of QOL and applying it within clinical practice has been a controversial issue for a number of years. The primary contentions have been directed towards the identification, selection and use of psychometric instruments in determining the level of morbidity in specific conditions, and the sensitivity of these

tools in detecting improvements in QOL after the appropriate therapy. Ideally, the particular psychological concerns of patients relating to a specific disorder should be assessed using accurate, disease-specific QOL questionnaires. However, practical experience reveals a large number of available generic QOL instruments that have been applied across many different disease states. Therefore, the implementation of such generic questionnaires may not provide a true reflection of the impact that a specific disease state may have on an individual's psychological well-being.

There are a number of generic questionnaires that have been applied to patients with GHD. These include the Nottingham Health Profile (NHP) (5), the Psychological General Well-Being Schedule (6) and the Rand 36-item Health Survey (SF-36) (7). These QOL instruments are divided into subcategories that can measure response in different areas of distress, though they do not specifically address the concerns of particular importance to patients with hypopituitarism. Consequently, their use has led to a large variation in results and their interpretation with respect to the effects of GHD and GH replacement therapy on QOL (8, 9). However, large-scale studies have since demonstrated that patients with GHD have a lower level of perceived QOL compared with a normative population (10), and that QOL in adults with GHD can be improved during and after GH replacement therapy (11, 12).

The most extensively validated GH-specific questionnaire developed to date is the Quality of Life – Assessment of Growth Hormone Deficiency in Adults (QOL-AGHDA) questionnaire (13, 14). However, though it is a self-rated questionnaire specifically tailored to assess QOL in GH-deficient patients, it does not consider that each individual will place a different level of importance on each aspect of their functioning, often referred to as 'items'. This can be achieved using a weighting scale for each item.

In response to the demand for a more disease-specific instrument to assess QOL in patients with hypopituitarism, we have developed and validated an additional disease-specific module of the Questions on Life Satisfaction^{Modules} questionnaire (QLS^M; initially constructed and tested in Germany as the FLZ^M) assessment (15). The QLS^M is a highly sensitive, subjective and multidimensional questionnaire for determining QOL. It comprises two other modules – overall life satisfaction (QLS^M-A) and general quality of health (QLS^M-G). The additional disease-specific module, QLS^M-H (Fig. 1), addresses the particular concerns of patients with hypopituitarism. A key feature of this modular questionnaire, distinguishing it from other disease-specific, self-rating questionnaires, is that each item is weighted according to its relative importance to the individual, with all three modules being concise and thus making it relatively easier for individuals to complete.

A second objective behind the development of the QLS^M-H was to establish a questionnaire able to evaluate therapy measures carried out as part of clinical studies or even in individual patients. Because such studies are increasingly internationally organised, the brief was to develop a questionnaire that could be applied globally across many different cultural backgrounds, while still taking the recommended methodological standards into consideration (16–18). After the formulation of the QLS^M-H, a series of investigations were undertaken to validate the questionnaire, primarily in patients with GHD and in normative population samples.

Patients and methods

The QLS^M-H was developed after studying relevant literature, followed by interviews with patients, in order to generate items. As a result, a preliminary German language version of the questionnaire was developed. This version was subsequently translated into five other languages, with studies being conducted in seven countries. On the basis of data generated from these studies, item selection was then carried out in a parallel approach. Subsequent psychometric evaluation was conducted individually in each country.

Interviews

As a basis for item generation, 26 semi-structured interviews were performed with patients with known, pre-existing hypopituitarism for more than 1 year. The patient cohort consisted of 19 men and seven women aged 24–64 years. Four patients were classified as having childhood onset (CO) GHD and 22 patients were defined as having adult onset (AO) GHD. Interview topics were based upon the patients' perception of specific aspects of life that may be influenced by their GHD disorder. These include physical fitness, ability to concentrate, relationships with friends and family, and feelings/mood. The interviews were conducted in the presence of two clinically experienced endocrinologists and lasted between 20 and 40 min. A significant number of categories demonstrating restricted QOL were reported, of which 16 items were included in the preliminary German version.

Translation procedure

The 16 items, together with the interview instructions and response categories, were translated from German into Dutch, UK English, US English, Italian and Spanish, by two independent bilingual translators for each language, with the primary objective of creating semantic equivalence. Each translated version was then back-translated into German and tested for semantic equivalence by the test authors (P Herschbach and G Henrich). For each translated version, approximately

QLS^M Questions on Life Satisfaction ————— Hypopituitarism

The following aspects of health are worth considering particularly in people with hormone deficiency. As with the previous questions, please indicate how important the individual aspects are to you personally and how satisfied you have been with them.

Again, please answer all questions and think about how you have been feeling over the last 4 weeks and not how you feel at this precise moment.

Please begin by putting a cross to show how important each individual aspect has been for your health in recent weeks. Before you begin, please look at all the aspects first.

How important for you is (are) your ...	not important	somewhat important	important	very important	extremely important
1. resilience/ability to tolerate stress _____	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
2. body shape _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. self confidence _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. ability to become sexually aroused _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. concentration _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. physical stamina _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. initiative/drive _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. ability to cope with your own anger _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. ability to tolerate noise and disturbance _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please put a cross now to show how satisfied you have been with these same aspects in recent weeks.

How satisfied are you with your ...	dis- satisfied	somewhat dissatisfied	somewhat satisfied	satisfied	very satisfied
1. resilience/ability to tolerate stress _____	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
2. body shape _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. self confidence _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. ability to become sexually aroused _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. concentration _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. physical stamina _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. initiative/drive _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. ability to cope with your own anger _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. ability to tolerate noise and disturbance _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

© Eli Lilly, 2001

Figure 1 The disease-specific module of the Questions of Life Satisfaction questionnaire designed for patients with hypopituitarism.

one-third of the categories did not agree with the original semantics and therefore could be misunderstood. After consultation with the translators, each of the remaining items was again subjected to the procedure described above. The revised version was then presented to a small group of patients in each country in order to test for comprehension and acceptance.

Item analysis and psychometric testing

The translated questionnaire was used in open, non-controlled, multicentre, longitudinal studies each with a similar design. All studies were designed also to examine the safety and efficacy of human GH (hGH) replacement therapy in patients with hypopituitarism. The studies were conducted in Australia, Germany, Italy, the Netherlands, Spain, the UK and the USA. The UK English version was used in the Australian study.

At the time of enrolment into the study, patients were receiving stable substitution of thyroid hormones, sex steroids, cortisol and/or vasopressin as appropriate. The dose of hGH was increased stepwise over 1–3 months increasing to a maximal dose of 12.5 µg/kg per day depending on the occurrence of GH-related adverse events. The duration of treatment was either 7 months in the USA or 6 months in the remaining countries. The time points at which the QOL questionnaires were used are presented in Table 1. Determination of body fat was performed either by bioelectric impedance assessment (BIA; in the German, Dutch, Spanish and US studies) or by dual-energy X-ray absorptiometry (DEXA; in all other studies). Insulin-like growth factor (IGF)-I was measured centrally using an IGF binding protein-blocked RIA. All studies were approved by the local ethics committees, and written informed consent was obtained from all participants.

The updated version of the QLS^M-H was developed on the basis of factor and item analyses and psychometric

evaluation focusing on the questionnaire's reliability, validity and sensitivity. The item selection was performed on the basis of a previously defined criteria catalogue, which was used in parallel in all seven participating countries. In addition, the two pre-existing modules of the QLS^M, the general (QLS^M-A) and the health-related module (QLS^M-G) (15), were used to determine whether GH treatment effects can be demonstrated with more sensitivity with the new hormone-specific module (QLS^M-H).

In the German and Spanish studies, the questionnaire was also used in interviews of a representative, normal population sample, because 'deviation from the norm' was one of the item selectivity criteria. To achieve this objective, independent social science research institutes were briefed to include a representative and coincidental selection of households in the survey. This survey comprised 1819 normal, adult individuals in Germany and 881 in Spain.

In order to test the construct validity, commonly-used generic life satisfaction questionnaires were used – the NHP (5) and the SF-36 (6). Table 1 provides an overview of the test design.

The test sample comprised 717 patients (at baseline) and 2700 normal persons. In three of the countries the design included an additional QOL assessment 4 weeks before the beginning of the GH replacement therapy, in order to evaluate test–retest reliability. The validation scales – NHP, SF-36 – were answered only by the patients, and not by individuals in the normal population.

Statistical analysis

Statistical evaluation was performed using the Statistical Package for Social Sciences (SPSS) program. Specific statistical methods that were used are described in the appropriate sections of the text.

Table 1 Overview of the test design. The hGH dose in the patient population was increased from month 1 to month 3, to a maximal dose of 12.5 µg/kg per day, depending on the presence of GH-related adverse events. Duration of treatment was either 7 months in the USA or 6 months in the remaining countries. The time points at which the appropriate QOL questionnaire (QLS^M, NHP or SF-36) was used are shown. *The time points in the USA were –1, 0, 4 and 7 months.

Country	No. of patients				Normative data (No.)	Validation scale
	Time point (month)					
	–1	0	3*	6*		
Australia	N/A	68	66	64	N/A	SF-36
Germany	151	146	136	126	1819	SF-36
Italy	N/A	113	97	93	N/A	SF-36
Netherlands	92	85	N/A	76	N/A	NHP
Spain	64	63	59	58	881	SF-36
UK	N/A	122	116	111	N/A	SF-36
USA	N/A	120	82	50	N/A	NHP
Total sample size	307	717	556	578	2700	

N/A, not applicable.

Table 2 The key demographic characteristics of the patient populations in the different countries.

	Australia	Germany	Italy	Netherlands	Spain	UK	USA
<i>n</i>	70	160	113	93	64	124	120
Age (years)							
Mean	44.4	44.1	41.2	42.1	39.6	45.8	48.6
S.D.	14.9	14.3	16.0	15.3	13.8	13.3	13.4
Sex (%)							
Male	60.0	56.9	65.5	45.2	43.7	53.2	66.1
Female	40.0	43.1	34.5	54.8	56.3	46.8	33.9
Marital status (%)							
Single	31.9	33.1	50.4	35.5	40.6	29.0	26.8
Married	65.2	61.9	44.2	53.8	50.0	58.1	56.4
Other	2.9	5.0	5.4	10.7	9.4	12.9	16.8
Occupational status (%)							
White-collar work	45.6	34.3	21.4	29.3	30.4	34.7	55.1
Blue-collar work	13.2	31.3	25.9	13.0	28.6	22.3	17.0
Unemployed	14.7	6.3	31.3	44.6	28.6	14.0	4.1
Retired	26.5	28.1	21.4	13.0	12.5	28.9	23.8

Results

Sample characteristics

Table 2 illustrates the key demographic characteristics of the patient sample. The mean age in the different countries varied from 39.6 to 48.6 years, with the percentage of women being between 33.9 and 56.3%. About half of the patient population were married (44.2–65.2%) and in employment (42.3–72.1%).

Table 3 describes some of the main medical variables of the patient groups. The majority of patients were diagnosed with AO GHD (61.9–86.8%) and having multiple pituitary hormone deficiency (MPHD) (88.4–100.0%). The mean body mass index (BMI) was increased compared with that in the average population in all countries, and varied between 26.7 kg/m² in Italy and 30.1 kg/m² in the UK. The proportion of endogenous fat mass ranged, on average,

from 25.6 to 34.3%. Mean serum IGF-I ranged from 48.4 µg/litre in Spain to 87.2 µg/litre in the USA. The last three variables (from all countries) are further differentiated in Table 4 according to time of onset of GHD (COGHD compared with AOGHD) and type of pituitary deficiency (isolated GHD (IGHD) compared with MPHD). Significant differences were observed between time of onset of GHD, and BMI and IGF-I level in each subgroup. The relative amount of fat mass was distributed evenly in the time of onset groups and in the type of disorder groups (IGHD compared with MPHD).

Item selection

The 16 *items* of the preliminary version of the questionnaire are listed in Table 5. They were assessed according to the combination of importance and satisfaction, providing the weighted satisfaction (WS).

Table 3 The major medical variables of the patient population. The percentage of fat mass was determined by BIA in Germany, the Netherlands, Spain and the USA and by DEXA in Australia, Italy and the UK.

	Australia	Germany	Italy	Netherlands	Spain	UK	USA
Onset (%)							
Childhood	23.2	21.9	38.1	21.5	31.2	16.1	13.2
Adult	76.8	78.1	61.9	78.5	68.8	83.9	86.8
Deficiency (%)							
IGHD	11.6	9.4	8.0	1.1	0.0	4.0	4.7
MPHD	88.4	90.6	92.0	98.9	100.0	96.0	95.3
BMI (kg/m ²)							
Mean	27.9	27.2	26.7	27.9	28.5	30.1	29.5
S.D.	4.8	5.3	5.0	5.0	7.4	5.7	6.2
Fat mass (%)							
Mean	29.7	30.8	25.6	34.3	27.2	31.4	33.4
S.D.	11.0	12.8	12.0	12.2	10.6	10.4	10.6
IGF-I (µg/litre)							
Mean	71.1	67.3	64.8	57.6	48.4	73.5	75.1
S.D.	41.0	50.8	45.2	26.8	18.2	36.6	52.1

Table 4 The major medical variables of the entire patient population, differentiated in terms of time of onset of disease (AOGHD or COGHD) and type of pituitary hormone deficiency (IGHD or MPHGD).

	Time of onset of GHD		Deficiency	
	AOGHD	COGHD	IGHD	MPHD
<i>n</i>	468	154	38	584
BMI (kg/m ²)				
Mean	28.7	26.0	28.4	28.0
s.d.	5.2	6.2	6.3	5.6
Fat mass (%)				
Mean	30.8	30.0	33.4	30.5
s.d.	11.3	13.7	11.0	12.0
IGF-I (μg/litre)				
Mean	74.3	48.9	87.5	66.6
s.d.	42.5	41.8	45.6	43.2

The formula for this combination (15) is:

$$WS = [\text{importance} - 1] \times [2 \times \text{satisfaction} - 5]$$

The importance scale of the questionnaire is unipolar (i.e. ranging from 1 = not important, to 5 = extremely important), whereas the satisfaction scale is bipolar (i.e. ranging from 1 = dissatisfied, to 5 = very satisfied, with a zero point between the 2nd and 3rd categories). Therefore, we subtract a score of 1 from the importance ratings in order that 'unimportant' answers are not considered in the total score of the QLS (multiplication with zero). The purpose of the formula '2 × satisfaction - 5' is to transform the zero

Table 6 Item selection: definition of psychometric terms.

Principal component analysis (PCA): A multivariate statistical technique that, essentially, reduces a correlation matrix into a few major pieces. It describes which variables 'go together' and aims at extracting principal components to condense as much of the total variation in the data as possible with a minimum number of components.

Oblique rotation: A non-orthogonal procedure, involving the rotation of one or more axes, so that the resulting factors will be correlated with each other.

Eigen value: The proportion of variance in a variable that is explained by the components.

Scree test: The concept behind the scree test is that, if a factor is important, it will have a large variance. The factors are ordered by variance and the variance is plotted against the factor number. The factors that lie above the 'elbow' in the plot are retained. These are the important factors that account for the bulk of the correlations in the matrix.

Factor loading: In factor analysis, this describes the correlation between one of the variables and the factor.

Selectivity, rit (correlation [r] of item [i] with test [t]): Part-whole correlation between item and scale, which differentiates individuals with a high occurrence of the construct from persons with a low occurrence of that construct.

Kolmogorov–Smirnov test: A statistical test that determines whether the distribution of a variable is different from the normal distribution.

Homogeneity (Cronbach's alpha): A measure of the consistency of a test scale; it examines whether all items of a scale measure the same construct.

Table 5 Factor loadings of the 16 items of the preliminary version of the QLS^M-H. The items shown in **bold** typeface are those chosen for the final version of the questionnaire.

Factors	Items	Factor loadings		
		Patients (<i>n</i> = 653)	Norm (D) (<i>n</i> = 1472)	Norm (E) (<i>n</i> = 782)
1	1 Resilience/ability to tolerate stress	0.94	0.96	0.96
2	2 Body shape	0.92	0.93	0.92
	3 Weight	0.94	0.93	0.91
3	4 Height	1.00	0.99	0.98
4	5 Sleep	0.97	0.96	0.96
5	6 Self confidence	0.60	0.70	0.75
	7 Social contact	0.97	0.93	0.92
6	8 Ability to become sexually aroused	1.00	0.99	0.98
7	9 Memory	0.95	0.94	0.93
	10 Concentration	0.92	0.93	0.92
8	11 Physical stamina	0.85	0.83	0.76
	12 Initiative/drive	0.90	0.90	0.85
9	13 Self control	0.89	0.91	0.75
	14 Cope with own anger	0.87	0.74	0.92
10	15 Ability to get peace and quiet	0.64	0.82	0.51
	16 Ability to tolerate noise and disturbance	0.83	0.86	0.88
	PCA: explained variance (%)	88.6	87.1	87.0

Norm, normal population; D, Germany; E, Spain.

point of the satisfaction scale to 2.5 (between the 2nd and 3rd categories).

The aim of the item selection was to establish a final set of items that contained, for economic reasons, minimal redundancy and which also complied with the principal psychometric quality criteria (see below). In order to minimize redundancy in the list of 16 items, similar items were grouped by exploratory principal component analyses (PCA) (Table 6) (19) with subsequent oblique rotation (Table 6). These analyses were performed separately in the patient population and in both normative samples. As a result of the PCAs, 10 factors were selected. This selection was based on the plot of the eigen values against the number of factors (scree test) (Table 6) (20). The assignment of items to the factors was performed according to the highest factor loading (>0.60) (see columns 1–3 in Table 5). The 10 factors explain 87.0–88.6% of the variance in the three samples.

The selection of factors and items within the factors was performed according to the psychometric criteria described below:

- (1) Sensitivity to change (*t* value of the *t*-test for dependent samples between month 0 and month 6).
- (2) Deviation from the normal population (*t*-test for independent samples between patients and the normative data).
- (3) Selectivity (rit; Table 6) of the items (correlation between item and scale).
- (4) Test–retest reliability (product–moment correlation between months –1 and 0).
- (5) Deviation from the normal distribution (*P* value of the Kolmogorov–Smirnov test (Table 6)).

The relevant analyses were performed for each of the 16 items separated in the samples from the seven countries. The evaluation of the difference between patients and the normative samples could be performed only in Germany and Spain, and test–retest reliability could be assessed in Germany, the Netherlands and Spain.

The result of these analyses was the final item selection consisting of eight items (shown in bold in Table 5): resilience/ability to tolerate stress, body shape, self-confidence, ability to become sexually aroused, concentration, physical stamina, ability to cope with own anger and ability to cope with noise and disturbance. A ninth item ('initiative/drive') shows a high loading on the factor that also contains item 11 ('physical stamina'; between 0.85 and 0.90). According to the established criteria, this item would not have

Table 7 The principal characteristics of the nine selected items that make up the final, updated QLS^M-H. The difficulty (mean), standard deviation (s.d.) and selectivity (rit) are shown.

Item	Population group						
	Australia	Germany	Italy	Netherlands	Spain	UK	USA
Resilience/ability to tolerate stress							
Mean (s.d.)	2.9 (6.7)	0.0 (5.9)	4.8 (7.1)	1.9 (5.3)	2.7 (5.9)	1.1 (6.7)	6.1 (8.2)
rit	0.71	0.69	0.63	0.76	0.68	0.68	0.70
Body shape							
Mean (s.d.)	–1.8 (6.6)	0.0 (5.4)	4.0 (6.8)	0.3 (6.6)	1.2 (6.7)	–2.6 (6.3)	0.6 (7.5)
rit	0.46	0.44	0.59	0.52	0.54	0.42	0.59
Self-confidence							
Mean (s.d.)	3.5 (6.9)	4.8 (7.2)	8.9 (7.7)	4.7 (7.8)	6.9 (8.0)	2.6 (7.7)	6.7 (7.9)
rit	0.59	0.60	0.65	0.68	0.74	0.69	0.70
Ability to become sexually aroused							
Mean (s.d.)	3.0 (6.4)	3.1 (6.1)	6.5 (8.6)	2.9 (6.0)	1.9 (7.6)	0.8 (6.8)	5.1 (7.8)
rit	0.20	0.49	0.34	0.48	0.51	0.40	0.41
Concentration							
Mean (s.d.)	4.1 (7.0)	2.3 (7.2)	7.2 (7.2)	4.3 (8.2)	3.4 (7.3)	1.3 (7.1)	5.4 (9.0)
rit	0.52	0.70	0.61	0.72	0.63	0.63	0.65
Physical stamina							
Mean (s.d.)	–2.0 (6.0)	–0.5 (6.0)	4.4 (8.5)	–0.2 (7.4)	0.4 (6.8)	–1.3 (6.9)	0.7 (8.2)
rit	0.51	0.66	0.76	0.69	0.62	0.57	0.67
Initiative/drive							
Mean (s.d.)	2.4 (6.6)	2.8 (7.4)	6.2 (6.8)	3.2 (7.1)	4.7 (6.7)	0.5 (6.6)	4.0 (8.0)
rit	0.64	0.71	0.50	0.77	0.66	0.60	0.76
Cope with own anger							
Mean (s.d.)	3.5 (6.8)	3.1 (6.8)	3.3 (6.2)	2.4 (6.2)	1.9 (7.4)	2.8 (7.8)	6.2 (8.2)
rit	0.38	0.60	0.51	0.65	0.71	0.58	0.65
Ability to tolerate noise/disturbance							
Mean (s.d.)	5.0 (5.9)	2.3 (6.8)	3.5 (6.8)	2.0 (6.8)	2.2 (6.3)	3.1 (7.0)	6.0 (7.7)
rit	0.55	0.57	0.40	0.61	0.66	0.69	0.69

been included in the final version of the questionnaire. However, clinical experience and considerations regarding the indicator function of 'initiative/drive', particularly for depression, led us to include this item also.

The main characteristics of each item, difficulty (mean), standard deviation (s.d.) and selectivity (rit), are presented in Table 7. The item scores can, theoretically, vary between -12 and +20. However, the actual average scores range between -2.6 (e.g. 'figure/appearance', UK) and +8.9 (e.g. 'self-confidence', Italy). The standard deviations vary between 5.3 and 9.0. The selectivity coefficients (rit) are excellent throughout (scores >0.4). Exceptions to this are the scores for 'sexual arousal' in Italy and Australia, and 'coping with anger' in Australia. However, these exceptions do not require the elimination of these items, as a certain dependence of results on the samples is acceptable in this context.

The psychometric properties of the scale

The scores from the nine items were added together to yield a total score. The mean scores from each country and the reliability of the scale are shown in Table 8. The mean total scores from different countries varied quite significantly. Patients in the UK study had the lowest life satisfaction, with a value of 7.7, whereas Italians had the highest score, at 49.4.

As an indication of the reliability of the scale, the homogeneity (Cronbach's alpha) (Table 6) and the test-retest reliability were assessed. The range of scores for Cronbach's alpha (0.81-0.89) indicate the good homogeneity of the scale. The test-retest reliability could be assessed only in Germany, the Netherlands and Spain. The baseline data from each study were

correlated over a single month without intervention. Correlation coefficients were 0.88, 0.87 and 0.76 respectively, indicating excellent stability.

The sensitivity to change in score as a result of GH replacement was assessed by the *t*-test for dependent samples, analysing the scores at baseline (month 0) and after 6 months (7 months in the US study). Figure 2 shows the mean scores of the QLS^M-H module for all countries before and after treatment. Table 8 illustrates the means of paired differences and the *t* value, together with the corresponding level of statistical significance. All scores increased during GH replacement therapy and changes proved to be highly significant. The changes were clearly greater than those observed from each of the non-specific modules of the QLS^M (QLS^M-A and QLS^M-G) as described elsewhere.

Internationally validated QOL scales were used for construct validation. Because no GHD-specific instruments existed at the time the data were collected, the NHP (5) and SF-36 (Medical Outcome Study Health Survey Short Form) (7) were used, as both scales have, in the past, been used extensively to assess QOL in adult patients with GHD. Table 9 shows the product-moment correlation coefficients of the QLS^M-H scale with the SF-36 (physical and mental component summary score) together with the six scales of the NHP. All correlations were significant. With respect to the SF-36, the correlation coefficients for the mental component summary score are, without exception, greater than for the physical component summary score. In the various NHP subscales, 'emotional reactions', 'social isolation' and 'energy level' exhibit the closest correlation to the QLS^M-H scale. The correlations are inverse - high scores in the NHP mean low QOL and *vice versa* for the QLS^M and SF-36 scales. Taken together, all these

Table 8 The mean total scores derived from the nine-item questionnaire for each country. Reliability of the scale was assessed with Cronbach's alpha representing homogeneity, together with test-retest reliability. The sensitivity to change in response to GH replacement therapy is also shown, together with the difference between the patient group and the normative population. All scores increased during GH replacement therapy and changes were significant.

	Population group						
	Australia	Germany	Italy	Netherlands	Spain	UK	USA
Total score							
Mean (s.d.)	20.9 (37.3)	23.8 (41.6)	49.4 (43.1)	27.6 (45.4)	24.6 (49.4)	7.7 (43.1)	41.1 (52.8)
Reliability							
Cronbach's alpha	0.81	0.87	0.84	0.89	0.89	0.86	0.89
Test-retest reliability		0.88		0.76	0.87		
Sensitivity to change (month 6-month 0)							
Mean of paired differences (s.d.)	30.1 (41.8)	22.4 (35.7)	14.5 (39.5)	12.1 (33.0)	25.7 (50.6)	40.4 (47.5)	25.2 (51.5)
<i>t</i> value	5.76***	6.87***	3.49***	3.18**	3.87***	8.79***	3.35**
Difference patients (month 0)- normative data							
Mean of paired differences (s.d.)		23.1 (50.3)			21.0 (68.1)		
<i>t</i> value		5.35***			2.24*		

****P* ≤ 0.001; ***P* ≤ 0.01; **P* ≤ 0.05.

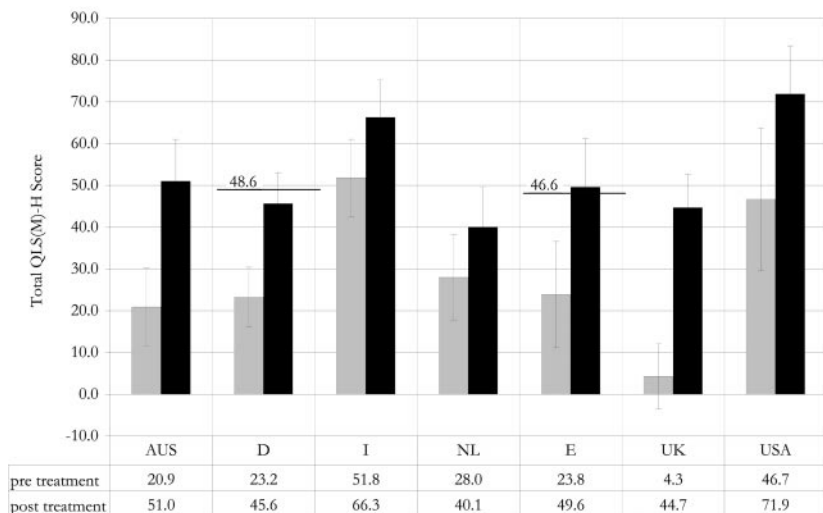


Figure 2 Total scores of the QLS^M-H scale before (pre; shaded bars) and after (post; dark bars) treatment in the seven countries. Also included are the normative scores (horizontal bars) for Germany (D) and Spain (E). AUS, Australia; I, Italy; NL, Netherlands.

data support the construct validity of the QLS^M-H scale.

A further validity criterion is the concurrent validity, which can be tested empirically by comparing the scores derived from the questionnaire given to objectively different populations (e.g. mildly and severely ill patients with the same diagnosis, or ill and healthy persons). For patients with hypopituitarism, there is no generally acknowledged single indicator of the severity of disease. We analysed the variables of disease onset (COGHD compared with AOGHD), number of deficient pituitary hormone axes (IGHD compared with MPHD), BMI, percent fat mass and serum IGF-I concentrations. These variables for each country were correlated with the total score from the QLS^M-H scale. However, no significant correlations were found.

We also tested the magnitude by which patients with hypopituitarism differ from the normal population when matched for age and sex. Table 8 demonstrates that, before treatment (month 0), the

patients had significantly lower scores and thus poorer life satisfaction than the normal population. After 6 months of replacement therapy, this difference had completely vanished.

Discussion

This report describes the development and psychometric testing of a novel QOL questionnaire for adult patients with hypopituitarism. Such an instrument must fulfil the following criteria in order to become a useful tool for clinical studies and diagnostics in individual patients. Firstly, it should be disease-specific. Secondly, it should permit a subjective weighting of single items. Thirdly, it should meet the classic psychometric quality criteria of validity, reliability and sensitivity to change. Fourthly, it should reveal deviation from the normal population, thereby having the potential to detect pathology. Fifthly, it should be globally applicable across different cultural

Table 9 The product-moment correlation coefficients of the QLS^M-H scale with the SF-36 (Medical Outcome Study Health Survey Short Form, physical and mental component summary score) and the six scales of the Nottingham Health Profile (NHP). The correlations with the NHP are inverse, because of the opposite 'polarity' of the two scales (high scores indicate good QOL in the QLS-H and poor QOL in the NHP and *vice versa*).

	Population group						
	Australia	Germany	Italy	Netherlands	Spain	UK	USA
SF-36							
Physical component summary	0.33	0.37	0.22		0.38	0.49	
Mental component summary	0.72	0.64	0.67		0.76	0.70	
NHP							
Physical mobility				-0.44			-0.50
Pain				-0.35			-0.47
Sleep				-0.28			-0.36
Social isolation				-0.56			-0.55
Emotional reactions				-0.72			-0.50
Energy level				-0.61			-0.43

backgrounds. Finally, it should be short but comprehensive, and well accepted by the patients.

In the past, the development of internationally applicable QOL questionnaires has been approached in a sequential, parallel or simultaneous fashion (16, 17). The primary objective of these three approaches is to control for significant cultural influences. Here, we used a procedure that may best be described as a parallel approach. A preliminary German language version of the questionnaire, complete with 16 items, was established on the basis of comprehensive interviews and published literature and then translated by independent translators into various languages. Back-translation into the original language was performed in order to ensure semantic equivalence to the highest possible degree. The 16-item version of the questionnaire was subsequently applied to patients from Australia, Germany, Italy, the Netherlands, Spain, the UK and the USA. In the German and Spanish studies, the questionnaire was also applied to normal population samples. Item reduction was carried out on the basis of the data obtained with these various samples. The predefined selection criteria were inter-correlation of the items or factor structure, sensitivity to change (sensitivity to treatment), deviation from the normal population, selectivity, test–retest reliability and distribution of the scores. The final updated version of the questionnaire comprises two identical sets of nine items – one asking the patient for the subjective importance of an item or the weight, with the other set asking the patient for his/her satisfaction with respect to that item. Thus, with this approach, the weighted satisfaction of single items can be assessed in addition to overall weighted satisfaction (sum of the individual scores).

The scores for Cronbach's alpha and test–retest correlations indicate good reliability, with the test of paired differences between the assessments before and after therapy indicating good sensitivity of the scale in all countries. Unsurprisingly, both the decrease in QOL at baseline and sensitivity to change after GH replacement therapy observed with the disease-specific instrument were clearly greater than that observed with the non-specific QLS^M modules (QLS^M-A and QLS^M-G). This indicates that QOL instruments that focus on a patient's specific disease-related problems are likely to be more accurate with respect to detecting pathology and the efficacy of subsequent treatment.

Assessment of the construct's validity in each country provided satisfactory results. The concurrent validity could also be supported, though only with regard to discriminating patients with GHD from the normal population. Within the patient group in each country, none of the reported disease variables showed consistent and significant mean score differences. This result is surprising, particularly with regard to the time of onset of disease. However, in the total patient sample, individuals with COGHD had better life

satisfaction than those patients with AOGHD. Whether this difference is due to the generally younger age of patients with COGHD – young, normal individuals also have a better life satisfaction than older, normal individuals – or whether it reflects different natural histories of the disorder, clearly remains to be established.

Clear variations between the different countries were evident at baseline. The total scores varied from 7.7 for the UK to 49.4 for Italy. To explain these differences, various factors may be considered to have an impact. Firstly, semantic equivalence of the questionnaires may not have been achieved, despite a stringent and standardized translation procedure. Secondly, patient selection may have significantly differed between the countries, though the inclusion criteria and, in particular, the definition of GHD, were similar between the various studies. Finally, it is possible that true cultural differences exist with respect to subjective life satisfaction. In support of the final suggestion, it has been established that differences in QOL between countries can be measured even among representative populations (21, 22), though these differences cannot be attributed to objective parameters (e.g. gross national product).

In contrast, the changes in scores during hGH replacement therapy were comparable in all the countries involved. Therefore, the QLS^M-H questionnaire appears to be a useful tool, at least in the longitudinal monitoring of GH replacement independently of cultural background. However, to exploit the power of this instrument fully, particularly with regard to detecting deviation from the norm, normal ranges need to be established in all countries. These findings do, however, highlight an important consideration: the pooling of QOL data from international studies is not a trivial matter, and must be approached with special caution.

The development of the GHD-specific QOL questionnaire that we describe here was conducted using several open-label, non-controlled hGH intervention studies. Ideally, a double-blind placebo-controlled study with an identical design in all countries would have been the most favourable approach. However, such a design was not possible after the approval of hGH treatment for adults with GHD. The majority of ethics review boards would have rejected such an approach in a comparable size of patient population. There is no doubt that QOL measurements can be influenced by the patient's expectations regarding therapy. However, the only variable that should be expected to be influenced by such expectations is sensitivity to change. The other variables investigated in this study such as test–retest reliability or deviation from the normal population refer to the baseline and should therefore be unaffected.

In conclusion, we have demonstrated the development of a new, specific instrument to assess QOL in patients with hypopituitarism, which can be applied

internationally and has been comprehensively tested by psychometric methods. It has a number of advantages over existing scales, allowing an individual weighting of single QOL aspects. The questionnaire is short, and thus economic, is well accepted by the patients and is suitable to be used not only in clinical studies but also for individual patients. Future evaluations in independent cohorts, ideally in placebo-controlled studies, will indicate whether this novel tool proves to be a powerful instrument in the evaluation of QOL in adult patients with GHD.

Acknowledgements

This study was sponsored by Eli Lilly & Company. The authors acknowledge, with gratitude, the contributions, time and effort of the investigators who were involved in collecting patients' data.

References

- Cuneo RC, Salomon F, McGauley GA & Sönksen PH. The growth hormone deficiency syndrome in adults. *Clinical Endocrinology* 1992 **37** 387–397.
- Bengtsson B-Å, Brummer R-J & Bosaeus I. Growth hormone and body composition. *Hormone Research* 1990 **33** (Suppl 4) 19–24.
- Deijen JB, de Boer H, Blok GJ & van der Veen EA. Cognitive impairments and mood disturbances in growth hormone deficient men. *Psychoneuroendocrinology* 1996 **21** 313–322.
- McGauley GA. Quality of life assessment before and after growth hormone treatment in adults with growth hormone deficiency. *Acta Paediatrica Scandinavica Supplement* 1989 **356** 70–72; discussion 73–74.
- Hunt SM, McEwan J & McKenna SP. *Measuring Health Status*. London: Croom Helm, 1986.
- Dupuy HJ. The psychological general well-being index. In *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*, pp 170–183. Eds NK Wenger, ME Mattson, CD Furberg & I Elinson. New York: Le Jacq Publications, 1984.
- Ware JE Jr & Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992 **30** 473–483.
- Whitehead HM, Boreham C, McIlrath EM, Sheridan B, Kennedy L, Atkinson AB *et al*. Growth hormone treatment of adults with growth hormone deficiency: results of a 13-month placebo controlled cross-over study. *Clinical Endocrinology* 1992 **36** 45–52.
- Rosén T, Wirén L, Wilhelmsen L, Wiklund I & Bengtsson B-Å. Decreased psychological well-being in adult patients with growth hormone deficiency. *Clinical Endocrinology* 1994 **40** 111–116.
- Cuneo RC, Judd S, Wallace JD, Perry-Keene D, Burger H, Lim-Tio S *et al*. The Australian multicenter trial of growth hormone (GH) treatment in GH-deficient adults. *Journal of Clinical Endocrinology and Metabolism* 1998 **83** 107–116.
- Wirén L, Bengtsson B-Å & Johannsson G. Beneficial effects of long-term GH replacement therapy on quality of life in adults with GH deficiency. *Clinical Endocrinology* 1998 **48** 613–620.
- Wallymahmed ME, Baker GA, Humphris G, Dewey M & MacFarlane IA. The development, reliability and validity of a disease specific quality of life model for adults with growth hormone deficiency. *Clinical Endocrinology* 1996 **44** 403–411.
- Holmes SJ, McKenna SP, Doward LC, Hunt SM & Shalet SM. Development of a questionnaire to assess the quality of life of adults with growth hormone deficiency. *Endocrinology and Metabolism* 1995 **2** 63–69.
- McKenna SP, Doward LC, Alonso J, Kohlmann T, Niero M, Prieto L *et al*. The QoL-AGHDA: an instrument for the assessment of quality of life in adults with growth hormone deficiency. *Quality of Life Research* 1999 **4** 373–383.
- Henrich G & Herschbach P. Questions on Life Satisfaction (FLZ^M) – a short questionnaire for assessing subjective quality of life. *European Journal of Psychological Assessment* 2000 **16** 150–159.
- Bullinger M, Anderson R, Cella D & Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality of Life Research* 1993 **2** 451–459.
- Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I & McKenna S. Cross-cultural adaptation of health measures. European Group for Health Management and Quality of Life Assessment. *Health Policy* 1991 **19** 33–44.
- Guillemin F, Bombardier C & Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993 **46** 1417–1432.
- Gorsuch RL. *Factor Analysis* (2nd edition). Hillsdale, New Jersey: Erlbaum, 1983.
- Cattell RB. The scree test for the number of factors. *Multivariate Behavioral Research* 1966 **1** 245–276.
- Hradil S & Immerfall S. *Die westeuropäischen Gesellschaften im Vergleich*. Opladen: Leske & Budrich, 1997.
- Arrindell WA, Hatzichristou C, Wensink J, Rosenberg E, van Twillert B, Stedema J *et al*. Dimensions of national culture as predictors of cross-national differences in subjective Well-Being. *Personality and Individual Differences* 1997 **23** 37–53.

Received 19 December 2000

Accepted 4 May 2001