Does long-term GH replacement therapy in hypopituitary adults with GH deficiency normalise quality of life?

Maria Kołtowska-Häggström1,2, Anders F Mattsson1, John P Monson3, Paul Kind4, Xavier Badia5, Felipe F Casanueva6, Jan Busschbach7, Hans P F Koppeschaar8 and Gudmundur Johannsson9

1KIGS/KIMS/ACROSTUDY Medical Outcomes, Endocrine Care, Pfizer, Sollentuna, Sweden, 2Department of Pharmacy, Uppsala University, Uppsala, Sweden, 3Department of Endocrinology, St Bartholomew’s Hospital, Queen Mary University of London, London, UK, 4Outcomes Research Group, Centre for Health Economics, University of York, York, UK, 5Health Outcomes Research Europe, Barcelona, Spain, 6Department of Medicine, Endocrine Section (FFC), School of Medicine and Complexo Hospitalario, University of Santiago de Compostela, Santiago de Compostela, Spain, 7Department for Medical Psychology and Psychotherapy, Erasmus, MC, Rotterdam, The Netherlands, 8Division of Endocrinology, Utrecht Academy, Medical Centre, Utrecht, The Netherlands and 9Department of Endocrinology, Sahlgrenska University Hospital at Gothenburg’s University, Gothenburg, Sweden

(Correspondence should be addressed to M Kołtowska-Häggström; Email: maria.koltowska-haggstrom@pfizer.com)

Abstract

Objective: To determine whether impaired quality of life (QoL) in adults with GH deficiency (GHD) is reversible with long-term GH therapy and whether the responses in QoL dimensions differ from each other.

Methods: QoL was measured by the Quality of Life–Assessment for Growth Hormone Deficiency in Adults (QoL-AGHDA) in general population samples in England & Wales, The Netherlands, Spain and Sweden (n = 892, 1038, 868 and 1682 respectively) and compared with corresponding patients’ data from KIMS (Pfizer International Metabolic Database) (n = 758, 247, 197 and 484 respectively) for 4–6 years a follow-up. The subsets of patients from England and Wales, and Sweden with longitudinal data for 5 years’ follow-up were also analysed. The change of the total QoL-AGHDA scores and responses within dimensions were evaluated. Subanalyses were performed to identify any specificity in response pattern for gender, age, disease-onset and aetiology.

Results: Irrespective of the degree of impairment, overall QoL improved dramatically in the first 12 months, with steady progress thereafter towards the country-specific population mean. Problems with memory and tiredness were the most serious burden for untreated patients, followed by tenseness, self-confidence and problems with socialising. With treatment, these improved in the reverse order, normalising for the latter three.

Conclusions: Long-term GH replacement results in sustained improvements towards the normative country-specific values in overall QoL and in most impaired dimensions. The lasting improvement and almost identical pattern of response in each patient subgroup and independent of the level of QoL impairment support the hypothesis that GHD may cause these patients’ psychological problems.
depression, remained significantly impaired. However, to the best of our knowledge, there are only a few studies on QoL in these patients in relation to the country-specific normative data (20, 21), possibly because of the paucity of such QoL reference data. Given the known cross-country differences (22), using country-specific normative data as a reference would be both a logical and a well-established way of estimating the QoL deficit in patients.

Only a few studies have reported the effect of GH replacement on the specific QoL domains particularly affected by GHD (5, 9, 17, 19, 23). Hence, our main research question was whether or not the impairment of QoL in hypopituitary adults with GHD is reversible with long-term GH replacement therapy and if so, whether there is any specificity in the QoL responses.

Subjects and methods

QoL measure

QoL was measured by the Quality of Life-Assessment for Growth Hormone Deficiency in Adults (QoL-AGHDA), a disease-oriented instrument based on the concept that QoL is the degree to which human needs are satisfied (24). It was developed following in-depth interviews of adult patients with GHD attending the Christie Hospital in Manchester, and consists of 25 items that evoke yes/no answers, acknowledging or denying certain problems (25). A high QoL-AGHDA score denotes poor QoL. The instrument is reliable and valid with a high level of internal consistency (26).

Subjects

Countries: population and patients. Population and patient data from four countries (England and Wales, The Netherlands, Spain and Sweden) were treated as separate study subgroups. The population data in each country were collected by separate surveys. The population data from England and Wales (28) were collected following a postal survey of 1190 individuals from a general population-survey panel maintained by the Outcomes Research Group in York. The data for the Dutch study were sampled from the market research database (TNS NIPO, The Dutch Institute for the Public Opinion and Market Research), which holds a panel of over 200,000 respondents interviewed regularly through the Internet. In July 2005, a nationally representative sample (1400 individuals) was drawn from the Dutch population aged over 18. The Spanish normative values for QoL-AGHDA published in 1998 by Badia et al. (29) were used in our study. A survey was performed in a random sample (n=1930) of the Barcelona general population. The questionnaires were administered by trained interviewers and self-completed by respondents. In Sweden, a questionnaire was sent out in 2004 to a random sample (n=3005) drawn from the population permanently registered in Sweden, aged 18–85 years.

The patients’ data from respective countries were retrieved from KIMS (Pfizer International Metabolic Database) (27). KIMS was launched in 1994 at the request of endocrinologists and healthcare decision makers to monitor the outcome and safety of long-term GH replacement therapy in those patients being treated in a conventional clinical setting (27). Till date, it contains data on over 11,000 adult patients with GHD worldwide. Data are collected at clinic visits on specially designed case report forms. The data collection process is monitored by study monitors. It is a condition of entry to KIMS that each centre obtains approval from its local ethics committee and that patients give informed consent, either verbally or in writing, depending on the local legal requirements. The respective country-specific cohorts were retrieved from the total database. Four subgroups of adult hypopituitary patients with GHD, aged 20–79 years (758, 247, 197 and 484 from England and Wales, The Netherlands, Spain and Sweden respectively), who had QoL-AGHDA scores at baseline were retrieved from the KIMS database. Owing to the nature of the database, which is an open, observational, non-interventional study, with an ongoing enrolment of new patients, the duration of follow-up varied among the countries from a maximum of 4 years in Spain to 8 years in Sweden. The Dutch patients were followed up to 6 years and the patients from England and Wales for 7 years. For the same reason, the number of patients at each time point decreased. QoL-AGHDA scores were obtained at baseline, and then annually (Fig. 1).

Statistical methods

Each country-specific set of data was treated as composite time-series, because of the varying times of follow-up.

For each country, and for each follow-up visit in a cross-sectional analysis, regression models were fitted to estimate the differences in mean total QoL-AGHDA score as well as in mean scores within the dimension of QoL-AGHDA between the general population and the KIMS patient population.

Total QoL-AGHDA score differences between the general population and KIMS patients. As the first step, a linear regression model was developed using total QoL-AGHDA scores for the general population and patients at entry into KIMS. The independent variables in the model were (i) type of a study cohort (general population, 1: KIMS patients, 2), (ii) age at visit (years), (iii) gender (0, male; 1, female) and (iv) interaction terms between these variables. Setting a significance level at
10% for entry to the regression model, its final structure was found to be:

Mean QoL-AGHDA score = f(pop, age, gender, pop*gender).

Age, gender and the interaction terms can be perceived as adjustment factors.

For each country and for each yearly visit, the differences in mean QoL-AGHDA scores between the general populations and KIMS patients were estimated using this model. The estimated differences were relatively constant over age (for interaction P > 0.10), even if, as was the case in most countries, the absolute mean QoL-AGHDA scores differed by age. Results were averaged for gender.

Analyses of patient subgroups. Analyses were also performed to identify any specificity within the patient subgroups. The pattern of response was checked for gender and age as described above, but also for disease-onset (childhood- vs adult-onset of pituitary disorder) and aetiology (non-functioning pituitary adenoma vs craniopharyngioma). These analyses were performed for each country separately as well as for the pooled data.

Special regression analyses were performed to check whether the calendar-year of entry into the study or exit from follow-up might influence the results. The underlying hypothesis was that the inclusion criteria associated with the studied outcome might have changed over time and that those patients who exited might not have responded in a similar way as those who continued treatment.

Also subsets of patients from England and Wales (77 patients) and Sweden (121 patients) with a 5-years’ longitudinal follow-up and complete (non-missing) QoL-AGHDA data at each time point were assembled and analysed.

For estimating the yearly change in a longitudinal analysis of mean QoL-AGHDA scores in the patient population, the repeated measurement regression was applied. The variance–covariance matrix was estimated as unstructured. Adjustment was made for age and gender.

QoL-AGHDA scores within dimensions clustering. A potential of five dimensions for clustering individual QoL-AGHDA items was indicated based on impairments in QoL specific for GHD. To check the overall importance of each of these dimensions, five visual analogue scales were designed and data on them collected in a parallel
study (authors’ unpublished data). The responses to individual items within QoL-AGHDA were coded as 0/1 for No/Yes, respectively. For clustering, the regression analysis was performed using the results from those five visual analogue scales as the dependent variables and the expected set of individual QoL-AGHDA items as the independent variables. This approach yielded dimensions for all QoL-AGHDA items, as follows: problems with memory and concentration, tiredness, tenseness, social isolation and self-confidence (Table 1).

**Differences between the general populations and KIMS patients.** As, the within-patient dimension scores are expected to be correlated, mixed-linear regression was used to analyse the mean difference by QoL-AGHDA dimensions between the general population and KIMS patients. The number of items differed by dimension, so that standardisation was necessary and this was achieved by computing the percentage of items within a dimension that a subject expressed problems with. The covariance structure was estimated with an unstructured variance–covariance matrix.

**Table 1 QoL-AGHDA dimensions.**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>QoL-AGHDA item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory and concentration</td>
<td>I have to struggle to finish jobs</td>
</tr>
<tr>
<td></td>
<td>I have to read things several times before they sink in</td>
</tr>
<tr>
<td></td>
<td>I often lose track of what I want to say</td>
</tr>
<tr>
<td></td>
<td>I often forget what people have said to me</td>
</tr>
<tr>
<td></td>
<td>I find it difficult to plan ahead</td>
</tr>
<tr>
<td></td>
<td>My memory lets me down</td>
</tr>
<tr>
<td>Tiredness</td>
<td>I feel a strong need to sleep during the day</td>
</tr>
<tr>
<td></td>
<td>It takes a lot of effort for me to do simple tasks</td>
</tr>
<tr>
<td></td>
<td>I have to push myself to things</td>
</tr>
<tr>
<td></td>
<td>I feel worn out even when I’ve not done anything</td>
</tr>
<tr>
<td></td>
<td>I often feel too tired to do the things I ought to do</td>
</tr>
<tr>
<td></td>
<td>I have to force myself to do all the things that need doing</td>
</tr>
<tr>
<td>Tenseness</td>
<td>I often have to force myself to stay awake</td>
</tr>
<tr>
<td></td>
<td>I have difficulty controlling my emotions</td>
</tr>
<tr>
<td></td>
<td>I often feel very tense</td>
</tr>
<tr>
<td>Social isolation</td>
<td>There are times when I feel very low</td>
</tr>
<tr>
<td></td>
<td>I often feel lonely even when I am with other people</td>
</tr>
<tr>
<td></td>
<td>It is difficult for me to make friends</td>
</tr>
<tr>
<td></td>
<td>I find it hard to mix with people</td>
</tr>
<tr>
<td></td>
<td>I avoid mixing with people I don’t know well</td>
</tr>
<tr>
<td></td>
<td>I am easily irritated by other people*</td>
</tr>
<tr>
<td>Self-confidence</td>
<td>I lack confidence</td>
</tr>
<tr>
<td></td>
<td>I feel as if I let people down</td>
</tr>
<tr>
<td></td>
<td>I avoid responsibilities if possible</td>
</tr>
<tr>
<td></td>
<td>I feel as if I’m a burden to people</td>
</tr>
</tbody>
</table>

*The item: “I am easily irritated by other people” fits marginally better into social isolation, although it could also be categorised within the dimension tenseness.

Adjustments were made for age and sex. Analyses were performed at yearly visits (cross-sectionally). As these statistics for dimensions had relatively lower precision and lower stability – especially at the later stages when there were few patients – the presented time-series was shown for rather fewer years of follow-up compared with the overall QoL-AGHDA score.

**General comments.** Significance levels were set to 5%. Ninety-five percent confidence intervals (CI) were of the Wald type. Analyses were performed using statistical analyses systems (SAS) software version 8.2. SAS procedure for General Linear Models (PROC GLM) was used for cross-sectional analyses of total QoL-AGHDA scores, and the mixed procedure (PROC MIXED) was used for longitudinal analyses of QoL-AGHDA total scores and for dimensional analyses. Missing values were treated as missing at random.

**Results**

**General population**

**Data from England and Wales.** The overall response rate was 84% but complete QoL-AGHDA questionnaires were obtained from 74% of respondents (n = 892; 56% of them women). The mean age (years) of the participants was 52.8 (S.E.M. 0.5), 55.3 (S.E.M. 0.7) for men and 50.9 (S.E.M. 0.6) for women. The mean (S.E.M.) for QoL-AGHDA scores was 6.2 (0.3) and 7.1 (0.3) in men and women respectively. The mean (S.E.M.) for the overall score (Fig. 1) was 6.7 (0.2) as previously published (28).

**Data from The Netherlands.** A total of 1075 responses were obtained (response rate, 77%), but the Dutch complete QoL-AGHDA data were available for 1038 individuals (mean age 47.2 years, S.E.M. 0.5) with almost equal gender distribution. The mean age (years) of men (n = 510) was 47.3 (S.E.M. 0.7) and of women (n = 528) 47.2 (S.E.M. 0.65). The mean (S.E.M.) of QoL-AGHDA score was 4.5 (0.2) and 5.3 (0.2) in men and women respectively. The mean (S.E.M.) of overall score (Fig. 1) was 4.9 (0.2).

**Data from Spain.** A total of 963 individuals agreed to participate (response rate 50%). The data for QoL-AGHDA were available for 868 individuals (57% women). The mean age (years) of the total sample was 46.3 (S.E.M. 0.6), of men 46.2 (S.E.M. 0.7) and of women 46.5 (S.E.M. 0.9). The mean (S.E.M.) QoL-AGHDA scores were 5.7 (0.2), 3.9 (0.2) and 5.0 (0.2) in men, women and the total sample, respectively (Fig. 1) (29).
**Data from Sweden.** In total, 1945 responses were received (response rate, 65%). A QoL-AGHDA score was available for 1682 (56%) respondents. The gender distribution in the sample was almost equal. The mean age (years) of the total sample was 48.3 (S.E.M. 0.4) with men being on average slightly older, 48.9 (S.E.M. 0.5), than women, 47.7 (S.E.M. 0.5). The mean (S.E.M.) QoL-AGHDA scores were 3.5 (0.2), 4.2 (0.2) and 3.8 (0.1) in men, women and the total sample, respectively (Fig. 1).

**Adult hypopituitary patients with GHD**

All patients had severe GHD confirmed by relevant stimulation tests and were naïve (had never received GH replacement before entry into KIMS). The main patients’ characteristics are presented in Table 2. In each country, GH and three additional pituitary hormone deficits were the most common combination (in patients coming from England and Wales, The Netherlands, Spain and Sweden were 37, 46, 47 and 41.5% respectively). All patients received appropriate hormonal replacement in relation to other pituitary hormone deficits. Surgery with or without radiotherapy was most commonly performed in patients from England and Wales (73%), The Netherlands (72%), followed by the Swedish (66%) and Spanish (57%) cohorts. The information on radiotherapy is included in Table 2. The most common co-morbidities reported at entry into KIMS were hypertension, arthritis and diabetes mellitus. Overall, 26% of patients had a history of fractures.

The mean duration of patient follow-up (years) was 3.2 for patients from England and Wales, 3.4 for the Dutch cohort, 2.5 for the Spanish and 4.5 for the Swedish.

**Cross-sectional, country-specific analyses of the composite time-series QoL-AGHDA overall scores**

Although at baseline, patients with GHD presented significant impairment in QoL (significantly higher QoL-scores), as compared with the respective normative data (Table 3), the greatest difference in the mean QoL-AGHDA score was found for England and Wales, followed by the Spanish, Dutch and Swedish patients. However, this difference diminished with time and led towards normalisation by the end of follow-up. In all countries, the main improvement occurred during the first year of treatment, followed by less rapid, but steady further normalisation (Fig. 1). There was a striking similarity between the treatment-response patterns among the countries; even though for Spain the data were available only for up to 4 years’ treatment.

The results for men and women were analysed separately for each country, demonstrating the same improvement curves (data not shown). The differences in total QoL-AGHDA scores (pooled data) between men (mean 11.0; 95% CI 10.6–11.4) and women (mean 13.2; 95% CI 12.6–13.6) were larger in the patient population (P > 0.0001) than in the general population (5.3 (5.1–5.6); 5.3 (5.0–5.5) (mean; 95% CI) for men and women respectively; P = NS). Additionally, the differences between patients and general population were greater in women but only at baseline. At later visits, as women responded to treatment significantly better than men (mean difference −0.8 (95% CI −1.6 to −0.1; P = 0.03)), both genders showed very similar differences compared with the general population. Overall, in the patient cohort women had higher QoL-AGHDA scores (poorer QoL) compared with men, but those differences diminished over the course of GH treatment.

### Table 2 KIMS patients’ clinical characteristics.

<table>
<thead>
<tr>
<th></th>
<th>England and Wales (n=758)</th>
<th>The Netherlands (n=247)</th>
<th>Spain (n=197)</th>
<th>Sweden (n=484)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender (n (%))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>363 (48)</td>
<td>123 (50)</td>
<td>75 (38)</td>
<td>247 (51)</td>
</tr>
<tr>
<td>Women</td>
<td>393 (52)</td>
<td>124 (50)</td>
<td>122 (62)</td>
<td>237 (49)</td>
</tr>
<tr>
<td><strong>Mean age at entry into KIMS (years; s.d.)</strong></td>
<td>48.5 (12.6)</td>
<td>48.2 (13.3)</td>
<td>45.0 (11.1)</td>
<td>51.4 (13.0)</td>
</tr>
<tr>
<td><strong>Disease onset (n (%))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Childhood-onset</td>
<td>54 (7)</td>
<td>25 (10)</td>
<td>21 (11)</td>
<td>31 (6)</td>
</tr>
<tr>
<td>Adult-onset</td>
<td>704 (93)</td>
<td>222 (90)</td>
<td>176 (89)</td>
<td>453 (94)</td>
</tr>
<tr>
<td><strong>Primary cause of GHD (n (%))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pituitary adenoma</td>
<td>510 (67)</td>
<td>156 (63)</td>
<td>81 (41)</td>
<td>314 (65)</td>
</tr>
<tr>
<td>Craniohypophyngioma</td>
<td>62 (8)</td>
<td>22 (9)</td>
<td>19 (9.5)</td>
<td>33 (7)</td>
</tr>
<tr>
<td>Other pituitary/hypothalamic tumours</td>
<td>44 (6)</td>
<td>13 (5)</td>
<td>9 (5)</td>
<td>21 (4)</td>
</tr>
<tr>
<td>Cranial tumours distant from pituitary/hypothalamus</td>
<td>24 (3)</td>
<td>7 (3)</td>
<td>4 (2)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Treatment for malignancy outside cranium</td>
<td>10 (1.5)</td>
<td>5 (2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other causes of acquired GHD</td>
<td>85 (11.5)</td>
<td>35 (14)</td>
<td>69 (35)</td>
<td>65 (13)</td>
</tr>
<tr>
<td>Idiopathic GHD</td>
<td>23 (3)</td>
<td>9 (4)</td>
<td>15 (7.5)</td>
<td>37 (8)</td>
</tr>
<tr>
<td>Panhypopituitarism (n (%))</td>
<td>133 (17)</td>
<td>40 (16)</td>
<td>44 (22)</td>
<td>61 (13)</td>
</tr>
<tr>
<td>Isolated GHD (n (%))</td>
<td>77 (10)</td>
<td>27 (11)</td>
<td>1 (0.5)</td>
<td>46 (9.5)</td>
</tr>
<tr>
<td>Irradiation with or without surgery (n (%))</td>
<td>396 (52)</td>
<td>109 (45)</td>
<td>49 (25)</td>
<td>175 (36)</td>
</tr>
</tbody>
</table>
Similar improvement curves were also observed in childhood- and adult-onset patients as well as in patients with non-functioning pituitary adenoma and craniopharyngioma as a primary aetiology, both for country-specific and pooled data (data not shown).

No statistically significant influence was found in regression analyses to check the influence of the calendar-year of entry on the results. The only deviation was for the Swedish patients commencing KIMS follow-up between 1998 and 2000, who seemed to improve their QoL during the first year of treatment less rapidly (data not shown). Similarly, there was no difference in pattern of response to GH between patients who continued follow-up in the database and those who terminated.

**Country-specific analyses of the QoL-AGHDA overall scores in patients with a 5-year longitudinal series of complete data at each time point**

For these subsets of patients, the estimated mean QoL-AGHDA scores at baseline (adjusted to 50 years of age) were 15.2 (s.e.m. 0.68) and 8.2 (s.e.m. 0.44) for the English and Welsh \((n=77)\) and Swedish \((n=121)\) patients, respectively. The differences between patient scores at baseline and the scores for the general population scores were \(-8.6\) (95\% CI \(-10.0\) to \(-7.2\); \(P<0.0001\)) for England and Wales and \(-4.0\) (95\% CI \(-4.9\) to \(-3.1\); \(P<0.0001\)) for Sweden.

These differences reduced markedly over the first year of treatment and remained subsequently not significantly different from those of the general population, for up to 5 years of treatment (Fig. 2). As for the total cohorts, the mean differences were relatively constant over age, i.e. being similar for each age category (\(P\) for test for interaction between age and populations at treatment start was \(0.13\) for England and Wales, and \(0.42\) for Sweden).

**Longitudinal analysis of change/trend over time for overall QoL-AGHDA score per country**

QoL-AGHDA score significantly improved during the first year of treatment (pooled data as well as country-specific data, \(P<0.0001\); Table 4). Although the comparisons between the estimates of change/year after 1 year to 3, 5 and 7 years of treatment showed no further statistically significant improvement, the initial positive change was maintained, leading towards the levels of QoL presented by respective populations (Table 4).

**Responses per country and dimension**

The analyses at baseline showed significant impairment in patients on all QoL dimensions compared with the general population scores at baseline and the scores for the general population (Table 4). Although the comparisons between the estimates of change/year after 1 year to 3, 5 and 7 years of treatment showed no further statistically significant improvement, the initial positive change was maintained, leading towards the levels of QoL presented by respective populations (Table 4).

**Figure 2** Country-specific differences in QoL-AGHDA total scores between general population and KIMS patients with complete QoL-AGHDA data during GH replacement.
respective population data. In all countries, problems with memory and tiredness greatly exceeded those reported in other dimensions and despite significant improvement during treatment these two dimensions remained adversely affected throughout the follow-up (Fig. 3). The first dimension to reach the population levels was social isolation regardless of the country of origin (Fig. 3), followed by tenseness and problems with self-confidence (Fig. 3). Similar analyses conducted separately in men and women (data not shown) confirmed the same pattern of responses.

Therefore, based on these results, the hypothetical order of dimensions by degree of impairment severity would be as follows: memory and concentration, tiredness, self-confidence, tenseness and social isolation. Strikingly, improvement with GH treatment occurred in the reverse order and interestingly, the greatest impairment did not improve the most rapidly (Fig. 4).

**Discussion**

We found that the long-term GH replacement in adults with GHD resulted in an improvement in overall QoL to levels almost identical to the country-specific normative values. The most dramatic improvement occurred during the first 12 months’ treatment, with a steady, but more gradual amelioration later leading eventually to approximate the population reference level. Moreover, this pattern of response occurred irrespective of the degree of QoL impairment. We also confirmed that problems with memory and tiredness were the most serious burden. The sequence of dimensional improvement – consistent across different patient groups – was a new finding, as was that changes in most dimensions (except memory and tiredness) brought scores up to population values.

Should a comparative population include either normal subjects or patients with other chronic disorders (3, 30, 31)? Or should it comprise unaffected siblings to control for genetic and social–environmental factors (32)? We believe that the appropriate comparators should be chosen according to the research question.

We studied large population samples assessing the QoL deficit in patients with GHD in relation to the normative values. The critical issue here is whether the population samples are both representative and comparable. In all countries, except Spain, data were collected by specialist agencies from the sample that covered geographically the whole country and subsequently weighted to reflect general population parameters. The data were collected over similar time periods and used comparable survey methods. However, the Dutch and English and Welsh samples were constituted from panels of individuals who agreed to
take part in surveys. This fact may explain the higher response rates in these countries.

Even so, we believe that the population samples are representative; though a few limitations related to the patient cohorts warrant further discussion. KIMS is an observational survey, and in principle such surveys reflect routine clinical practice with no restrictions on the number of patients enrolled, duration of follow-up or strict study inclusion criteria (33, 34). This explains the variation in the extent to which QoL was impaired in English and Welsh patients, compared with that elsewhere. In the UK, the major criterion for GH replacement is impaired QoL, so that levels of QoL in GH-replaced patients (KIMS cohort) might have been worse than in the GH-deficient population at large. Conversely, in Sweden all patients with severe GHD qualify for GH replacement, so that the KIMS sample would have reflected average Swedish adult patients with GHD.

Another feature of long-term observational surveys is the varying time of patients’ recruitment – in our case from 1996 to 2003. The result was a decrease in the number of patients over the duration of the study, which was caused not only by patient exit or drop-out as well as missing data, but to a major extent by asynchronous entry into the database. Again, in open observational
surveys, data may be missing at certain time points. The additional analyses to check for the potential confounders (patients entered at different calendar-years, patients who continued follow-up vs those who terminated subsets of patients with no missing data) all yielded, and similar results and thus, confirming the validity of our findings. Particularly reassuring were the results of analyses of the patient subgroups with 5 years’ continuous follow-up and complete data, which indicated similar improvement of the QoL-AGHDA scores compared with the whole cohort (the ‘composite’ series). These results confirmed that the assumption of missing data at random might be justified, suggesting that these were not associated with the total QoL-AGHDA score.

Despite the lack of rigid inclusion criteria, the patient cohorts from all the four countries seemed to have similar characteristics. To account for the other potential confounders, several comparisons were performed (males vs females, patients with craniopharyngioma vs those with pituitary adenoma). Additionally, all results were statistically adjusted for gender and age, thus mirroring the demographic profile. Surprisingly, all results from different subgroups confirmed the same response pattern with little variations in the way QoL improved. Moreover, despite worse QoL at baseline and a shorter follow-up, patients with craniopharyngioma showed a similar beneficial response.

In contrast to the results of Malik et al. (19), our findings indicate a strong convergence with the population values of QoL-AGHDA. Among several possible explanations for these conflicting findings, the duration of follow-up is probably critical. In our study, QoL in English and Welsh patients dramatically improved during the first year of treatment, but still reached the population values only after 6–7 years. On average, those patients whose total QoL-AGHDA results over 5 years’ follow-up were available reached the population level much earlier, although very few patients were in this subgroup, which was probably highly selective. Both our population and patient cohorts were much larger than those of Malik et al. and thus may convey a more comprehensive picture. Interestingly, the mean QoL-AGHDA score in their controls was approximately one and a half points lower than we found. Such a difference in reference values could affect the results, particularly concerning deficits. Another striking difference between the two series was the primary aetiology of GHD. Malik et al.’s patients with craniopharyngioma comprised 15% of the total cohort (19); in ours, the proportion was 8%. Such severe underlying pathology might partly explain the conflicting results, although recently, Verhelst et al. have shown that a worse QoL was found only in men with craniopharyngioma compared with those with non-functioning pituitary adenoma with the opposite being true for women (35). Our subgroup of patients with craniopharyngioma had slightly more profoundly impaired QoL compared with patients with non-functioning pituitary adenoma; however, their improvement was comparable.

The second part of our study examined the dimensions covered by QoL-AGHDA. Confirming the results of other studies, we found that the most extensive impairment before starting GH treatment...
was related to memory (6, 13, 14) and vitality (5–9). Nevertheless, all functions were reduced, social isolation being the least impaired. Our detailed analyses of long-term responses to GH treatment suggested that these are not only consistent but also valid for various patient groups. Attempting to identify the most sensitive response in psychological functioning with treatment, Burman suggested that these were energy and emotions (17). Using the Nottingham Health Profile (NHP) (energy levels, emotional reactions and physical mobility), Wirén found that up to 50 months of treatment, the response was very similar to ours; the main improvement occurred during the first 6 months, and less thereafter (36). Recently, Mukherjee confirmed these conclusions in a specific subgroup of patients with GHD – namely, cancer survivors (23). Using the Psychological General Well-Being (PGWB), she inferred that vitality was the most profoundly impaired before treatment, and improved the most with treatment. Interestingly, the pattern of response she found did not differ for any outcome in either cancer survivors or patients with other pituitary disease. This agrees with our findings.

Despite dramatic changes, particularly during the first year of treatment, memory and tiredness did not reach population levels during the whole observation period. The first dimension to stabilise within normative values was that least impaired – namely, social isolation – followed by tenseness and self-confidence. Such improvement occurred in a reverse order to the pretreatment severity. Nevertheless, all our results were measured by subjective indices. Possibly, measured by objective tests – for example, associate learning task or associated learning delayed recognition task (memory) or treadmill (energy) – both memory and energy might respond differently, as Arwert found for memory (14). Again, conceivably all five dimensions might constitute a form of pyramid, with memory and energy being at the base, and hence the closest function of strictly biological parameters and the most dependent on the pure somatic environment. Tenseness and self-confidence incorporate more of psychological factors and finally proper social functioning, despite its deep roots in each of them, depends on many other aspects. Hence, although still distant from population norms, a substantial improvement in memory and vitality (together with other factors) might be sufficient for normalisation within the more externally or superficially located dimensions.

In conclusion, long-term GH replacement resulted in dramatic and sustained improvement, leading towards the normative country-specific values in overall QoL, as well as in the dimensions reduced by GHD. The sustained nature of these ameliorations, and almost identical pattern of response independent of the level of QoL impairment, indicate that GHD per se may cause many of the psychological problems these patients face.

Acknowledgements

We thank the participants of KIMS in The Netherlands, Spain, Sweden and the UK who provided the data on their patients and the colleagues working at KICS/KIMS/ACROSTUDY Medical Outcomes in Stockholm. Without their commitment, support and daily hard work the KIMS database would not exist and many studies would have never happened. We also thank Dominique Maiter (Brussels, Belgium) for indicating the right direction towards the values that are important.

References

13. Baum HBA, Katzenelson L, Sherman JC, Biller BMK, Hayden DL, Schoenfeld DA, Cannistraro KE & Klibanski A. Effects of


33 Black N. What observational studies can offer decision makers? *Hormone Research* 1999 51 (Suppl 1) 44–49.

34 Kahn MG. Clinical research databases and clinical decision making in chronic diseases. *Hormone Research* 1999 51 (Suppl 1) 50–57.


Received 24 January 2006
Accepted 31 March 2006