University of Sheffield

ScHARR

SCHOOL OF HEALTH AND RELATED RESEARCH

RESPONSE TO COMMENTS RECEIVED FROM VARIOUS STAKEHOLDERS RESPONDING TO THE POST APPEAL CONSIDERATIONS FOR THE CLINICAL AND COST EFFECTIVENESS OF RECOMBINANT HUMAN GROWTH HORMONE (SOMATROPIN) IN ADULTS

PREPARED ON BEHALF OF THE NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE

FINAL DRAFT

20 January 2003

Nick Bansback
John Brazier
Jim Chilcott

Introduction

This report has been prepared following the responses to the post appeal considerations received from the following companies/organisations:

Eli Lilly
Novo Nordisk
Pharmacia
Pituitary foundation
Royal College of Physicians, Society for Endocrinology, BSPED & RCPCH

The key issues that were raised relate to:
1. Utility estimates reassessed using the EQ-5D
2. The selection of treatment defined by change in AGHDA and not baseline AGHDA
3. Effect of recombinant human growth hormone on mortality
4. The QLS and other quality of life instruments

We have not examined the transition from paediatric to adult care and survivors of childhood cancer.

1. Utility estimates reassessed using the EQ-5D

Background: *In Pharmacia's initial work, the item responses to the Qol-AGHDA questionnaire, used in the KIMS database, were translated to the Nottingham Health profile which in turn was converted to a quality of life single index via the SF-6D.(This utility term derived from the AGHDA score is referred to as the QoL-AGHDA utility) Whilst it would be preferable to have a direct utility measurement from the KIMS dataset, the conversion was reasonable in terms of the results that it gave, and the methods used. A new study from Pharmacia compares the utility score of patients using two methods, (1) the existing conversion from the AGHDA questionnaire and (2) the EQ-5D questionnaire. The results suggest that the actual utility improvement seen by the patients in the KIMs database may be larger than previously estimated.*

The two methods used for quantifying utility, the SF-6D and the EQ-5D, have well documented explanations for why one would expect to find differences between them. The SF-6D has a floor effect , restricting the patients utility score to above 0.3. In a comparison between the two instruments, it would be anticipated to see a greater degree of utility change using the EQ-5D in comparison to the QoL-AGHDA utility(via the SF-6D).

The report on the new valuation study involving 197 patients did not describe the results in sufficient detail to use in the ScHARR economic model. It is reported that 40% difference exists in the utility values estimated by the two instruments, but the actual association between the QoL-AGHDA utility and the EQ-5D is not reported.

We have re-estimated this equation using further data provided by Pharmacia and the resultant equations are presented for the EQ-5D and QoL-AGHDA utility in Figure 1. The EQ-5D model has a substantially lower R^2, but this is partly because the QoL-AGHDA utility is based directly on the AGHDA. The fits are significant overall, as is the difference between the two gradients.

The original mean AGHDA scores, used in the original submission, by age and baseline AGHDA are shown in Table 1. The conversion to QoL-AGHDA utility is

described in table 2 and conversion to EQ-5D is shown in Table 3. The differences in utility estimates vary depending on the age/AGHDA group, some are higher than 40% (max=55%) whilst some are lower.

Figure 1: A comparison between the Qol-AGHDA derived utility and EQ-5D derived utility

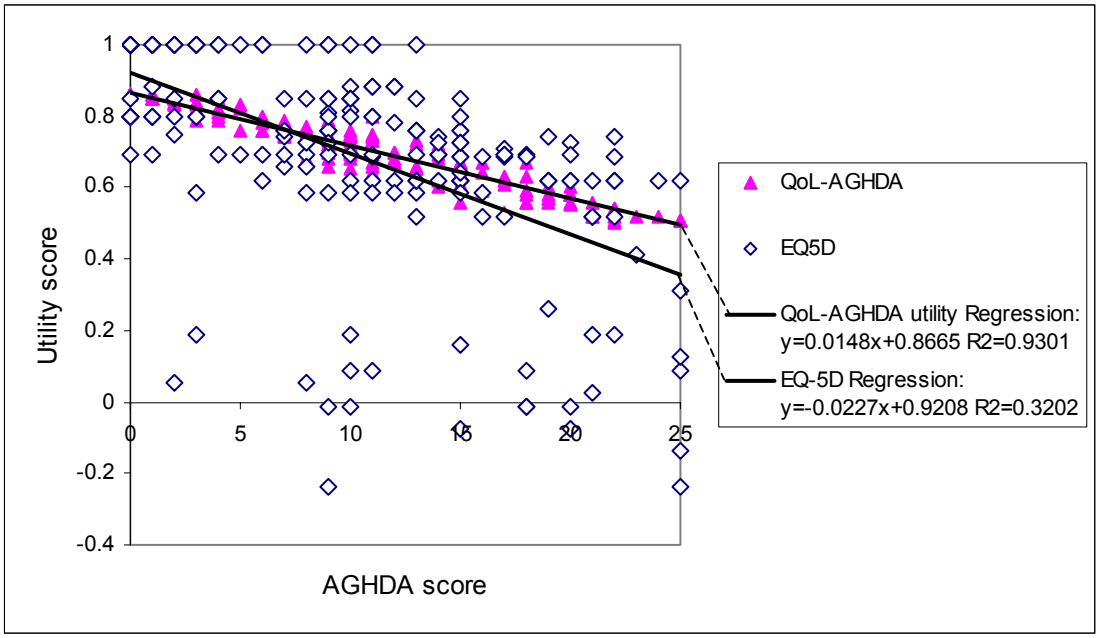


Table 1: AGHDA score at baseline and 1 year (from KIMS)

Time (Years)	< 65 years AGHDA score				> 65 years AGHDA score			
	0-5	6-10	11-15	16+	0-5	6-10	11-15	16+
0	2.09	8.04	13.08	19.52	1.95	8.11	12.74	20.07
1	1.80	5.55	7.75	11.98	1.94	3.82	5.22	13.17

Table 2: Conversion from AGHDA to QOL-AGHDA utility at baseline and 1 year (from KIMS)

Time (Years)	< 65 years AGHDA score				> 65 years AGHDA score			
	0-5	6-10	11-15	16+	0-5	6-10	11-15	16+
0	0.84	0.75	0.67	0.58	0.84	0.73	0.67	0.57
1	0.84	0.79	0.76	0.70	0.84	0.81	0.78	0.69

Table 3: Conversion from AGHDA to EQ-5D utility

Time (Years)	< 65 years AGHDA score				> 65 years AGHDA score			
	0-5	6-10	11-15	16+	0-5	6-10	11-15	16+
0	0.87	0.74	0.62	0.48	0.88	0.74	0.63	0.47
1	0.88	0.79	0.74	0.65	0.88	0.83	0.80	0.62

Table 4: Comparison of the utility change from QOL-AGHDA and EQ-5D

Time (Years)	< 65 years AGHDA score				> 65 years AGHDA score			
	0-5	6-10	11-15	16+	0-5	6-10	11-15	16+
Qol-AGHDA	0.00	0.04	0.09	0.12	0.00	0.08	0.11	0.12
EQ-5D	0.01	0.06	0.12	0.17	0.00	0.10	0.17	0.16
Difference	-	41.3%	34.4%	42.6%	-	21.7%	55.2%	30.5%

When these differences are included into the analysis the ICER are reduced. Tables 5 and 6 show the impact.

Table 5: Previous ICER estimates using the QOL-AGHDA

Qol-AGHDA group	Age Group				Overall by AGHDA
	18-30	31-55	56-64	65+	
0-5	*	*	*	*	
6-10	£124,941	£114,789	£94,866	£38,185	£94,123
11-15	£55,358	£50,884	£42,420	£27,885	£47,471
16+	£40,746	£37,483	£30,971	£25,286	£36,738
Overall	**£45,136**				

Table 6: Updated ICER estimates using the EQ-5D

Qol-AGHDA group	Age Group				Overall by AGHDA
	18-30	31-55	56-64	65+	
0-5					
6-10	£88,431	£81,304	£67,204	£31,484	£69,132
11-15	£41,182	£37,879	£31,513	£18,071	£34,729
16+	£28,573	£26,302	£21,733	£19,424	£25,819
Overall	**£32,210**				

Conclusion

The EQ-5D study shows that the ICERs derived from the ScHARR economic model, may underestimate the utility gain according to the AGHDA score to which it is derived. However, whilst we appear to endorse these revised estimates, a number of concerns exist over the methods used for acquiring the AGHDA/utility gains.

- The AGHDA changes are estimated from a single arm observational study with no comparator arm. The only RCT evidence showed significantly lower AGHDA gains.
- The AGHDA changes are calculated from baseline and year 1 estimates. The number of observations at year 1 is lower than at baseline, and therefore the estimates are open to selection bias.
- The EQ-5D study is based on single observations. Without a before and after study, one is not able to discern the genuine utility gains.

2. The selection of treatment defined by change in AGHDA and not baseline AGHDA

Background: *In the teleconference of the 26th November it was mentioned by a number of participants that the criteria for treating patients should not be based on the patients baseline quality of life (quantified via the AGHDA) but on the change in QOL the patient experiences in the initial period of treatment. Baseline AGHDA has been used as the criteria for selecting suitable patients in all current analysis since this was the approach used in the Pharmacia submission. The approach based on change in the initial 'trial' period seems to be a reasonable criteria for selecting patients, but the ICER of this decision is difficult to quantify without obtaining patient level data from the KIMS database. Data was not submitted by Pharmacia or KIMS on this matter. A crude estimate has been made to indicate the probability of patients in different Age/AGHDA groups, being in a cost effective region. This can be used to determine whether additional work with patient level data would be meaningful.*

In the initial economic analysis, the utility gains have been estimated from the KIMS database, and have been analysed by AGHDA and age groups. These changes in utility estimates provide us with the mean changes in quality of life of patients in each of those groups. Without the standard errors around the QOL changes, it is impossible to estimate the proportion of patients in each of the Age/AGHDA groups is achieving a QOL gain sufficient to be in a cost effective region. The impact of two possible criteria have been examined.

<u>Response based exclusively on a single AGHDA score improvement</u>

If the criteria for response is defined a single improvement in AGHDA score for all patients, and not based on their baseline age or AGHDA, then a simple threshold analysis shows that a 0.146 change in utility would be required for the overall ICER to be under £30,000. This can be converted to AGHDA via the two regressions shown in figure 1.

	ICER ratio		
	£30,000	£20,000	£40,000
Utility change	0.146	0.219	0.109
Converted to AGHDA via QOL-AGHDA utility	9.9	14.8	7.4
Converted to AGHDA via EQ-5D	6.4	9.6	4.8

To put these AGHDA figures into context, the greatest improvement in AGHDA score seen in the KIMS analysis was 7.5.

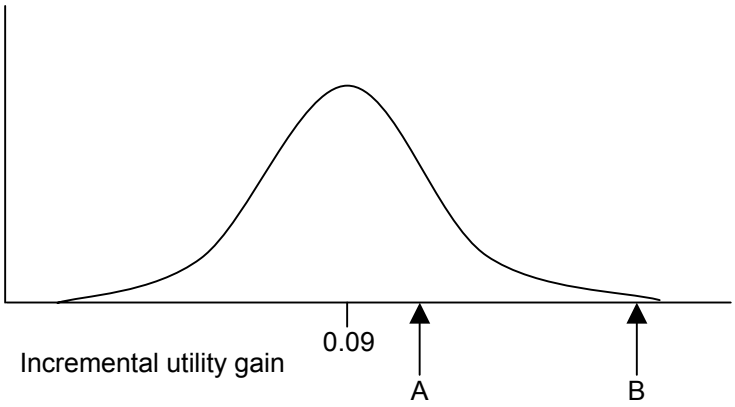
<u>Response based on baseline AGHDA score and AGHDA score improvement</u>

The key issue is to determine what proportion of patients in each of the age/AGHDA groups would achieve the utility gain required to be in cost effective region. The magnitude of this probability would increase in line with the baseline AGHDA. We do not know the proportion of patients from each group that will achieve this. I.e. if only 1% of patients in the 11-15 group achieve an AGHDA gain sufficient to meet the

required cost effectiveness then we would know that baseline AGHDA would be a suitable criteria for allocating treatment.

For example, patients in the AGHDA group 11-15, aged between 18 to 30 had a baseline utility score of 0.67 (KIMS). After 1 year this had risen to 0.76, therefore a mean utility change of 0.09. A threshold technique shows that patients in this group would need to achieve a utility gain of 0.142 to be considered cost effective(based on a ICER of £30,000) i.e. a 58% increase to the mean value. Without the distribution around the AGHDA gain seen in KIMs, it is impossible to calculate the proportion of the population that would meet the threshold. i.e. Figure 2 shows that it could be within the 95% CI (A) or not (B). (We do not even know whether the distribution is normally distributed)

Figure 2: An example of the uncertainty in the distribution around AGHDA gain



Conclusion

A simple threshold technique shows that an AGHDA score of between 6.4 and 9.9 (dependent on AGHDA/utility relationship) would be required for the entire age and baseline AGHDA mix of currently treated patients to be in a cost effectiveness region of under £30,000. A number of discrete factors exist between each of the age/baseline AGHDA groups making a single AGHDA gain for the criteria for continuation of treatment difficult. We do not have sufficient data to examine the uncertainty in AGHDA gain in each of the groups.

3. Effect of recombinant human growth hormone on mortality

Background: *The initial modelling work by Pharmacia estimated mortality benefit of rHGH by using both the Framingham risk equations and the Rotterdam fracture risk approximation. The Framingham Risk equations estimate the probabilities of CHD, MI, stroke and associated mortality based on a number of factors. The Rotterdam Risk equations estimate the probability of hip fracture in osteoporotic populations, and hence resulting mortality risk. The inputs to the equations, age, sex, total and HDH cholesterol, systolic blood pressure and bone mineral density were compared at baseline to 1 year after initiation of treatment. Of the subgroups that were analysed, the differences in mortality risk at baseline versus 1 year were very small. The resulting mortality benefit had a minimal impact on the ICER (~1%).*

In Pharmacia's response to the post appeal consideration, a new approach has been used to estimate the impact on mortality. The proposed rationale for changing the method of analysis is that the previous approach does not incorporate all mortality benefits identified with patients who are deficient in human growth hormone. Whilst this matter was not upheld in the appeal, it was deemed necessary by NICE to examine this further analysis.

The basic analysis estimates Standard Mortality Ratio (SMR) of an untreated growth hormone deficient population from the published literature. Five studies from various countries were selected, which described patients not treated with growth hormone replacement. These five studies provide SMR estimates between 1.6 and 2.1. A meta analysis by Pharmacia reports an average SMR figure of 1.87 for untreated patients. No overall confidence interval is reported. This figure is then compared to a SMR calculated from the KIMS database. This SMR for treated patients is given as 1.1, again with no confidence interval reported. The analysis then utilises the new modelling framework to quantify the impact of these new SMR estimates, in conjunction with the new utility estimates, to accrue the overall benefit of treatment in terms of QALYs. The impact of switching to this method of quantifying mortality impact, on the ICER is substantial (e.g. £18,900).

The most important problem with this method for calculating mortality difference is potential confounding. Without a trial of rHGH including a control arm of no treatment, evidence of a mortality benefit of rHGH is open to numerous confounding and bias factors, and therefore extremely weak. A comparison of SMRs across a number of single arm observational studies does not establish the differential mortality in the population of treated and non-treated patients, but rather shows the comparative mortalities in the trials patient groups. No account is made for any of the potential differences in co-morbidities, which may exist, between five trials and the KIMS database, let alone detailed analysis of survival curves, loss of follow-up etc.

When considering the rationale for accepting the new method, it also is important to examine its plausibility in both quantitative and clinical terms. If the substantial increase in estimated mortality impact is true, then it must be occurring via a mechanism, which is not measured by either the Framingham or Rotterdam risk factors. Since most mortality in human growth hormone deficiency is cardio-vascular, this implies that patients have some other unidentified risk factor substantially altered by treatment, which in turn reduces mortality. Pharmacia's response to the post appeal consideration does not discuss this issue and provides no plausible clinical mechanism for the mortality reduction.

These basic problems with this new mortality analysis are fundamental. The are also problems with the data used to construct the SMR based argument, which are recorded here for completeness. There was no systematic search used to identify the relevant sources for untreated patients, and no assessment of quality or suitability of the papers made. The selection of studies and data appears problematic. For instance one study mentioned in the original Pharmacia submission (Bates et al) shows an SMR of 1.2 (confidence interval 0.95 to 1.55) but is excluded from the meta analysis (with no reason given). Secondly, one study used in the meta analysis (Tomlinson et al) actually reports that only 11% of patients are tested for growth hormone deficiency, and that in this subgroup of the trial, there is no effect on mortality.

Conclusion

The revised analysis of mortality benefits substantially alters the estimated mortality impact and the subsequent cost-effectiveness calculations. It is based on observational data, with no adjustment for potential confounding. The calculations give very different answers from Framingham, implying a different clinical mechanism for mortality reduction, but with no plausible rationale discussed.

Without a controlled trial it is difficult to estimate the mortality benefit of rHGH. The most robust approach available to us, is using the before and after changes in mortality factors (cardiovascular and bone mineral density) estimated from the treated patient group from KIMs. This was the method used in the original assessment by Pharmacia, where mortality benefit had a minimal impact (~1%).

4. The QLS & other utility estimates

Background: *Eli Lilly have submitted further CIC documents relating to the results of work done using the QLS, a new quality of life instrument designed specifically for patients with growth hormone deficiency*

This subject has been discussed on a number of previous occasions. ScHARRs original stance referring to the information on the QLS was: *'it did not provide additional evidence meeting the inclusion criteria of our review of clinical effectiveness (e.g. a published RCT or longitudinal observational set) nor would the information have been able to contribute to the economic model since it is currently not possible to translate the QLS into utility values'*. Whilst this stance remains for its use in the economic model, the QLS could have a role in the future for use in patient selection.

The AGHDA was not selected due to its superiority in terms of deciding between different states of quality of life in rHGH patients or quality, but rather due its use in the most comprehensive model and availability in the most detailed data. More importantly is its ability to be converted to utility and so can be used to create cost utility estimates. If the use of rHGH is conditioned on baseline quality of life, this does not imply a restriction of the instrument used to necessarily assess quality of life. Specifically if future developments of the QLS or other instruments demonstrate that they can be validly converted to utility, then there is no reason for them not to be used. The QLS appears valid for this group from the available unpublished sources, and we understand a study is planned to validate it against utilities.