PERGAMON

European Journal of Cancer 36 (2000) 1459–1461

European
Journal of
Cancer

www.ejconline.com

Letters

# Comments on: *Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the UK*, Glaser AW, Furlong W, Waltzer DA, *et al.* *Eur J Cancer* 1999, **35**, 256–261

D. Dlugos *, B. Ravina, C. Jacobstein, A. Chimwaza, A. Siderowf

*Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 1st Floor Blockley Hall, Philadelphia, PA 19104, USA*

Received 15 February 2000; accepted 29 March 2000

Glaser and colleagues [1] report the use of the Health Utilities Index (HUI) to measure the health-related quality of life (HRQOL) in 30 children who had survived tumours of the central nervous system. The primary objective was to assess the applicability of the HUI (developed in Canada) to the measurement of health status in survivors of childhood cancer in the UK. Because of limitations in the psychometric data reported, we respectfully disagree with the authors' conclusions that acceptable reliability and validity of the HUI were established in this population.

The authors measured reliability by examining intra- and interobserver agreement. Intra-observer agreement, usually measured by test–retest reliability, was not reported for the HUI as a whole, but only for "...two identical questions about emotion." This is an unconvincing measure of the test–retest reliability of the entire instrument. Inter-observer agreement was reported for four different groups of raters (children, parents, physicians and physiotherapists) and is also unconvincing. Agreement for single attribute ratings were moderate or better (i.e. kappa greater than 0.40) for all raters only for the domain of self-care. For the domains of emotion, cognition and pain, agreement beyond chance was poor or slight, with kappas consistently not significantly different from zero. Confidence intervals were not reported, but given the small sample size, even the moderate correlations between raters are probably questionable. Correlation between raters for the global utility score was improved over the single attribute scores, but was still disturbingly low for the rater pairs physiotherapists/children (intraclass correlation coefficient — ICC = 0.40) and physicians/children (ICC = 0.15). The authors claim that they were "not surprised" by this result, point out that there is no gold standard for assessment of these items in children and that "...assessments provided by all types of assessors be respected." Whilst true, the possibility that the HUI may be a flawed instrument with poorly documented reliability in this population is not addressed.

There are also concerns about the validity of the HUI in this population. The authors report that construct validity was supported by "strong similarities" between the results obtained in their study from the UK and those found in a similar group of children in Canada [2]. However, significant differences were found between the two groups in the areas of emotion and self-care. Although the authors point out that the differences in the measurement of emotion between the two groups "may be due entirely to rewording of the questionnaire related to this attribute," this raises additional concerns about the implications of rewording an "established" instrument. The validity of the results may also have been affected by the age of the population. The authors collected data from children as young as 5 years who completed the questionnaire with the help of a play specialist. Even with the assistance of a play specialist, the youngest children may have been too young to comprehend fully the meaning of the questions.

A final area of concern involves suspected ceiling effects of the HUI. In this population, 21% of patients had normal health states or no deficit as ranked by the HUI. In another study evaluating adult stroke patients [3], nearly 50% of patients ranked themselves as 1 (the highest score) in each domain. Such a ceiling effect limits responsiveness and discriminative validity. Ceiling

---

* Corresponding author. Tel.: +1-215-590-4859; fax: +1-215-590-2950.

*E-mail address:* dlugos@email.chop.edu (D. Dlugos).

PII: S0959-8049(00)00124-6

effects may be particularly notable in certain populations, such as those with neurological disorders. Using sensation as an example, level 1 in HUI indicates "able to see, hear and speak normally for age," whilst level 2 "requires equipment to see or hear or speak." It is plausible that an aphasic or dysarthric subject could be rated as level 1 because he/she does not use an assistive device. Thus, the HUI may not be appropriate for populations with severe disabilities since floor effects are unlikely given the scale's design.

In summary, Glaser and colleagues reach overly optimistic conclusions about the utility of the HUI in paediatric brain tumour patients. The fundamental psychometric issues outlined above need to be addressed in more detail before the instrument can be recommended for use in clinical studies.

## References

1. Glaser AW, Furlong W, Walker DA, *et al*. Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the UK. *Eur J Cancer* 1999, **35**, 256–261.
2. Barr RD, Simpson T, Whitton A, Rush B, Furlong W, Feeny DH. Health-related quality of life in survivors of tumours of the central nervous system in childhood. *Eur J Cancer* 1999, **35**, 248–255.
3. Mathias SD, Bates MM, Patsa DJ, Cisternas MG, Feeny D, Patrick DL. Use of the Health Utilities Index with stroke patients and their caregivers. *Stroke* 1997, **28**, 1888–1894.

---

# Response from D. Feeny, *et al.*

D. Feeny[a], R. Barr[b], W. Furlong[c], A. Glaser[d,*]

[a]*Health Economics, University of Alberta, Canada*
[b]*Paediatrics, McMaster University, Canada*
[c]*Centre for Health Economics and Policy Analysis, McMaster University, Canada*
[d]*St James's University Hospital, Leeds, UK*

Received 10 April 2000; accepted 10 April 2000

We thank Dlugos and colleagues for raising important issues concerning the report by Glaser and associates [1] about the usefulness of the Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3) systems in the context of survivors of tumours of the central nervous system in childhood. We will deal with each of the points raised in their letter.

It is agreed that the assessment of test–retest reliability of the modified question used to determine HUI2 and HUI3 levels of emotion is not evidence of test–retest reliability for the HUI. Indeed Glaser and associates [1] 1999 indicate (p. 260) that the modification is 'inadvisable'. Evidence on test–retest reliability for HUI is found elsewhere [2–4]. Evidence on inter-rater reliability for HUI in this context is available also [2,5–8] whilst evidence on the construct validity of HUI in the context of brain tumours or other conditions involving cognitive impairment has been published [2,5,8–20].

Dlugos and colleagues do not state what level of agreement among assessors they expected to find. Perhaps they expected a high level of agreement. Yet abundant evidence for paediatric and adult measures of health-related quality of life (HRQL) indicates that substantial and systematic differences in opinions among assessors are not at all uncommon. Herjanic and Reich [21] administered simultaneously a structured diagnostic interview consisting of 168 questions to 307 children and their mothers. High agreement (kappa $\geqslant 0.50$) was observed for 16 (9.5%) of the questions; middle agreement (kappa 0.30–0.49) was observed for 30 (17.9%); low agreement (kappa $\leqslant 0.29$) was observed for 122 (72.6%). Many of the latter dealt with factors that are not readily observable, such as emotional health. We feel that our lack of surprise in not observing high agreement is well founded; evidence on a lack of agreement among assessors in paediatric settings using other instruments is substantial [22–25].

Dlugos and colleagues argue that aphasic or dysarthric subjects might be categorised as level 1 sensation in HUI2. Indeed it is possible, but it is very unlikely if the data were collected using sensation-related questions from the standard 15-item HUI2/3 questionnaire as in Glaser and associates [1]. It is unlikely because the level for HUI2 sensation is not determined directly from one question of the 15-item questionnaire, but rather from previously determined HUI3 levels of vision, hearing and speech. To be classified as level 1 sensation, a subject must have already been classified as level 1 vision and level 1 hearing and level 1 speech, based on combinations of responses to 6 questions. It seems highly unlikely that these types of subjects would be classified as level 1 speech in HUI3.

Dlugos and colleagues express concern about a suspected ceiling effect in that 21% of patients had no

---

* Corresponding author. Tel.: +44-113-206-4985; fax: +44-113-247-0248.

*E-mail address:* adam.glaser@gw.sjsuh.northy.nhs.uk (A. Glaser).