Quantifying Health Status Outcomes in Pediatric Medulloblastoma Patients

JOHANNES E.A. WOLFF1, ULRICH HÜTTERMANN2 and MARTHA A. ASKINS1

1Division of Pediatrics, University of Texas, U.S.A.; 2Division of Pediatrics, University of Münster, Germany

Abstract. Background: Comprehensive, efficient health status assessment tools are needed for multi-center studies examining childhood brain tumor treatment outcomes. The Fertigkeitenskala Münster Heidelberg (FMH) is presented as a quantitative measure of health status. Patients and Methods: The FMH was compared with the medical assessments, intelligence scores, and behavioral/emotional adjustment scores of 21 survivors of medulloblastoma to examine the instrument’s feasibility and discriminate validity. The subjects, aged from 3 to 38 years, represented a broad range of health functioning, from asymptomatic to severely handicapped. The median age at diagnosis was 10.5 years. The median time of assessment was 1.5 years from diagnosis. Results: Correlations calculated between the physicians’ assessments and the other measures showed a very close relationship between the physicians’ ratings and the FMH scores (rho=0.784, p<0.0005) and physicians’ ratings and Performance IQ (rho=0.679, p=0.005). Additionally, a correlation between the FMH and Full Scale IQ was demonstrated (r=0.627, p=0.012). The Child Behavior Checklist (CBCL) did not correlate with any other tests. Of the three assessment methods, the FMH showed the highest correlation with degree of handicap (p<0.0005). The FMH took less than 10 minutes to administer and was the only test feasible in all patients. Conclusion: The FMH is useful as an objective, easily administered measure of health status in brain tumor patients. The FMH offers health status information that is correlated with physicians’ assessments and FSIQ, but not behavioral/emotional adjustment as assessed by the CBCL.

Medulloblastoma is the most frequently occurring malignant brain tumor in childhood. Well-established treatment protocols include surgery, radiation and intensive chemotherapy (1, 2). Using these protocols, more than half of the patients will be long term survivors (3). This has created a growing population of medulloblastoma survivors and an increasing need to assess and address long term consequences of the disease and its treatment. Most of these consequences are negative (4-6), such as impaired cognitive capacity (7), sensory-perceptual capabilities and motor development. Neuro-cognitive decline across time is consistently found in visual-motor integration, visual memory, verbal fluency and executive functioning, while verbal memory and receptive vocabulary are often preserved (8, 9). In addition, attention problems have been associated with lower reading and math achievement performance in children with medulloblastoma (9). The contributions of younger age at treatment and higher cranial irradiation dose are the most critical to neuropsychological outcome in medulloblastoma (10, 11). Cognitive problems result in inferior school functioning and necessitate special education services (12). The overall health-related quality of life of children with brain tumors is highly variable, yet significantly lower than normal (7), with children who have received cranial spinal radiation faring worse overall. The impact of such late effects on the social and psychological adjustment of survivors of medulloblastoma and their quality of life is not yet well documented; however, Ribi and colleagues (12) reported that, in comparison with healthy controls, social functioning was rated by the patients as the quality of life dimension most affected.

For survivors of medulloblastoma, quality of life becomes a critical component of their therapeutic outcome. Quality of life is a broad concept, encompassing physical health and well-being, psychological health, educational or occupational capacity, social and family life, and standard of living (6). While each of these factors is important in the overall quality of life of the survivor, the health status is perhaps the most central and has a direct influence on many of the others. For this reason, determination of health status in the cancer survivor – and particularly in pediatric oncology patients – is essential in the measurement of treatment outcome and effectiveness.
Numerous approaches are available to assess health status and selecting the most appropriate approach depends on the goal of the assessment. Where improvement in treatment protocols is the primary goal, a global quantitative measure, suitable for multi-center studies is necessary. The German Brain Tumor Study Group is in the process of standardizing health status and quality of life measurements in multi-institutional settings (13, 14). In preparation for this, the Study Group tested the feasibility of a newly developed instrument for measuring health status in medulloblastoma patients – the Fertigkeitenkskala Münster Heidelberg (FMH) scale – comparing the results with tumor status, neurological exams, audiology, ophthalmology and endocrinology assessments. In addition, the results of the health status measure were compared to commonly utilized measures of other aspects of quality of life, including cognitive capacity and behavioral/emotional adjustment to determine discriminate (divergent) validity of the measure.

The group hypothesized that the FMH would provide a readily administered and unique measure of global health status and treatment outcome. This hypothesis was tested by: i) determining the number of individuals among our population of medulloblastoma patients to whom the instrument could be successfully administered; and ii) measuring the discriminative validity (utility) of the FMH in comparison to other frequently used instruments.

**Patients and Methods**

Between 1968 and 1995, 60 medulloblastoma patients were treated at the University Hospital of Münster. Prognostic factors, treatment and survival of the cohort have been previously published (15). Of the 28 survivors, 21 agreed to participate in this study. The male/female ratio was 10/11. The median age at diagnosis was 10.5 years (range=1.5 to 32.3 years). Patients were examined after a median time of 1.5 years from diagnosis (range=0.5 to 14 years), resulting in a mean age at testing of 11.1 years (range=3.10 to 38.7 years). All patients had radiotherapy as part of their tumor treatment, including 55 Gy posterior fossa dosage.

This study included a physical exam, endocrinology, audiology, and ophthalmology assessments, and an MRI of the head and spine. Following these assessments, the attending physician who reviewed these findings performed an interview with at least one parent and the patient. The physician then scored the sequelae according to predetermined criteria, as identified in Table I.

Subsequently, two additional measures of various aspects of functioning were administered where possible. The individuals conducting these assessments were unaware of the results of the previous assessments, or of the FMH assessment. The first of these, the Hamburg Wechsler Intelligence Scores for adults or for children (IQ), was used in the revised German version (16, 17). Second, the Child Behavior Checklist (CBCL) (18) was completed to examine the existence of behavioral/emotional problems.

Finally, the FMH assessment (13, 14) was completed on each of the patients. The FMH consists of 56 items (as outlined in Table II) designed to determine the health status of any individual. The items on this checklist were selected to describe functioning necessary for independence in daily life. Normalization studies on the scale have been conducted using randomly selected samples from a total population including both healthy and physically handicapped subjects, numerically weighted according to available population statistics. This means that approximately 9% of the data pool from which the normalization sample was selected were handicapped. Transforming the individual scores to percentiles had allowed an age independent description of health status. The value required for normal functioning is 50; values below 10 indicate severe reduction of independence in daily life as compared to peers.

The IQ tests were administered according to standardized procedures by trained examiners. Neither the CBCL nor the FMH requires training for administration. Both the CBCL and the FMH were completed by one parent of the patient.

**Results**

*Physical findings.* All patients were in complete remission at the time of the examination. In three patients, this was a second complete remission; it was the first complete remission in all others. Four patients were under active oncological therapy at the time of investigation. On physical exam, the most frequent neurological sign was ataxia (n=11), followed by nystagmus (n=10); paresis of the 7th and the 3rd cranial nerve was found in two patients each, paresis of the 6th cranial nerve in one. One patient was hemiplegic and one quadriplegic. There were five patients without any neurological symptoms. All patients had some degree of hair loss mainly within the radiation field of the
boost to the posterior fossa. On audiological examination, only one patient was free of any signs of impairment; three patients had unilateral hearing loss of high frequencies (higher than 4,000 Hz - NIH score I); 11 patients had bilateral hearing loss at high frequencies (NIH score I); one patient had unilateral hearing loss at lower frequencies (NIH II); two further patients had this finding in both ears and required hearing aids (NIH III). On ophthalmological exam, 6 patients were normal, 7 patients required glasses because of developed or increased myopia during their therapy. The remaining pathological ophthalmologic findings were related to cranial nerve involvement as mentioned above. On endocrinological exam, three patients had primary hypothyreosis, one patient had type I diabetes mellitus prior to oncological treatment. Various degrees of growth-related findings were present in most of the patients: 13 of 20 had dropped in body length percentiles during brain tumor treatment, 8 of 12 patients had pathologically low somatomedin-C (Insulin-like growth Factor I) levels in serum. One patient was treated with growth hormone after completing endocrinological testing.

Global physician’s assessments (Figure 1) were based on the physician’s knowledge of the above-listed physical findings, as well as on interviews with patients and parents. The patient’s performance on the remaining measures (IQ, FMH, and CBCL) was unknown to the classifying physicians. Five patients were described by the physician’s assessment as asymptomatic (neurological score: 5), six as mildly handicapped (neurological score: 4), 6 as significantly handicapped (neurological score: 3) and 4 as severely handicapped (neurological score: 2).

Intelligence testing (Figures 1, 2) could be conducted in only 15 patients. Three patients were too young, two had language difficulties and one refused. The mean IQ of the 15 patients was 92 with a range of 55-113 and a standard deviation of 14.6. The scores on the Verbal subtests were

**Table II. FMH Questionnaire to assess health status.**

<table>
<thead>
<tr>
<th>Locomotion</th>
<th>Read/Write/Calculate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can move both arms and both legs.</td>
<td>Distinguishes between one and many.</td>
</tr>
<tr>
<td>Can turn from prone to supine position.</td>
<td>Can count to three.</td>
</tr>
<tr>
<td>Can sit up without help and sit freely.</td>
<td>Understands a picture-book story.</td>
</tr>
<tr>
<td>Can drive the wheelchair without help.</td>
<td>Writes single words without copying.</td>
</tr>
<tr>
<td>Can walk stairs up and down holding onto something (banister, stick).</td>
<td>Can make calculations up to 100.</td>
</tr>
<tr>
<td>Can walk without holding onto something.</td>
<td>Reads simple books.</td>
</tr>
<tr>
<td>Can walk stairs up and down without holding onto something.</td>
<td>Can tell the date correctly.</td>
</tr>
<tr>
<td>Can ride a bicycle without training wheels.</td>
<td>Can write a short letter.</td>
</tr>
<tr>
<td>Can travel alone on bus or train.</td>
<td>Has published in the last year.</td>
</tr>
<tr>
<td>Has a driver’s license for a car or a motorbike.</td>
<td>Can calculate probabilities.</td>
</tr>
<tr>
<td>Can drive a car in a foreign city.</td>
<td>General Independence</td>
</tr>
<tr>
<td></td>
<td>Can see.</td>
</tr>
<tr>
<td></td>
<td>Can stay alone at least for one hour.</td>
</tr>
<tr>
<td></td>
<td>Can sleep alone in the dark with the door closed.</td>
</tr>
<tr>
<td></td>
<td>Knows his/her address.</td>
</tr>
<tr>
<td></td>
<td>Can go shopping by her/himself.</td>
</tr>
<tr>
<td></td>
<td>Can go alone to an administrative office.</td>
</tr>
<tr>
<td></td>
<td>Lives independent from parents/nursing care.</td>
</tr>
<tr>
<td></td>
<td>Earns his/her own money.</td>
</tr>
<tr>
<td></td>
<td>Provides leadership in a department / company.</td>
</tr>
</tbody>
</table>

Answers are “yes” or “no”. Yes answers are counted and transformed to age-dependent percentiles using a published table of normal values (7). Note that these items are culture dependent and cannot be transposed to an English language setting without redoing the normalization and validation of the questionnaire.

525
higher than Performance scores, with a mean of 102, a range of 65-119 and a standard deviation of 15.8. The Performance subtests had a median score of 82 with a range of 55-115, and a standard deviation of 16.9. There was no significant correlation between Performance and Verbal subtest scores (Table III).

The Child Behavior Checklist (CBCL) (18) was completed for only 14 patients. In addition to the reasons for incomplete testing described in the cognitive assessment, several parents found single items of this offensive and did not complete the test. Unfortunately, this is a common problem with the German-language version of the CBCL, in which a number of the questions can sound offensive. The mean percentile score of patients (parents) completing the test was 73 (range 33 to 98). High numbers on this test indicate symptoms in psychological, psychiatric, or social domains. The normal percentile value is 50. Percentiles 93 and higher are considered to be in the borderline clinical range, because they

Figure 1. Health status measurements in patient groups with various degrees of handicap: patients were classified according to the score described in Table I. Quantitative measures of the FMH (Table II), the Verbal and Performance scores of the Hamburg Wechsler intelligence test (Verbal and Performance IQ) and the CBCL are shown with average and standard errors. Note that the FMH results showed the most convincing relationship to the handicap score. Margin of error marks/error bars represent Standard Deviation.

Figure 2. Relationship between FMH and Performance IQ scores. The two measures were correlated with a Pearson correlation of 0.627, which was statistically significant (p=0.012 in two-tailed testing).
are high enough to be of concern, but no so high as to be clearly deviant. Percentiles of 97 and higher are felt to be of clinical significance (19). Clinically significant elevations occurred on the Internalizing and Externalizing scores in four patients. The only part of the test that never exceeded clinical significance was aggressive behavior. The most frequent item to be of symptomatic interest was item 24, which questioned eating behavior. Only half of our patients were normal for this item. After the CBCL was completed, a debriefing interview was held which addressed issues that came up with the test. Interestingly, neither the general judgement of the parents nor the general judgement of the interviewer confirmed the CBCL test results. Several of the single items suggested psychologically pathological symptoms, but were in fact caused by neurological or treatment-related (e.g. dysphagia) phenomena.

The FMH (Figures 1, 2) was completed in all 21 patients. The mean percentile score was 56 (standard deviation: 13.1). All five patients age 18 years or older scored 50% or higher. All patients scoring below 30% were younger than 10 years old.

Convergent and discriminant validity of the FMH. Non-parametric tests for correlation (Spearman’s p) calculated between the physician’s score and the other measures (Table IV) showed a very close relationship between the physician’s score and the FMH score (p=0.794, p<0.0005). A slightly smaller, but still statistically significant relationship was found between the physician’s score and the Performance score on the Wechsler intelligence test (p=0.679, p=0.005, Figure 2). Use of the statistically more powerful parametric test (Pearson’s r; Table III), further indicated a correlation between the FMH and the Full Scale IQ (r=0.627, p=0.012). However, there was no correlation between the Verbal IQ score on the intelligence test or the CBCL with that of the FMH. Other than the above-mentioned correlation with the physician’s score, Performance IQ was also unrelated to any other tests.

Discussion

The long-term sequelae of therapy for medulloblastoma in the present group of patients appeared slightly lower overall than those described in earlier studies (20-23). Many of these sequelae, such as motor symptoms like ataxia, reduced cranial nerve functions, and reduced cognitive processing speed have been reported to have an impact on the subsequent development of intelligence, academic performance and independence after completion of oncological treatment in pediatric brain tumor patients (22, 23). These phenomena can ultimately amplify differences in ability between brain tumor patients and their peers. The cause for these deficits in brain tumor patients is multifactorial. Long lasting hydrocephalus, neurosurgery, radiotherapy (especially in young children) and high dose methotrexate chemotherapy might contribute to this course. Sorting out the relative harm of each of these elements and developing treatment strategies resulting in fewer or less significant long-term sequelae is one of the major challenges in pediatric neuro-oncology. To achieve this, a sensitive, objective and achievable instrument to measure long term sequelae is necessary.

For it to be utilized in improving treatment protocols in multicenter studies, an instrument must have several properties: it obviously should measure the most important components of health status for this specific question. The measurement should be quantitative as opposed to qualitative, since results from various individuals have to be numerically combined and statistically tested for differences.
The instrument should be readily accepted by as many patients as possible and easy to use. Finally, the measurement cannot place an excessive burden on available resources. Overly-ambitious test batteries in previous studies at national, as well as at international levels have frequently resulted in minimal or no data coming in to study coordinators.

The intelligence tests have a recognized advantage of being well-standardized, well-known and widely-accepted. They measure some aspects of health status and quality of life with high sensitivity. This was confirmed by our data. However, our populations included severely handicapped patients with almost normal IQ. These instruments do not measure certain other aspects of health status and long-term sequelae at all. A survivor, who has a severe hemiparesis after tumor surgery and is completely deaf after cisplatinum chemotherapy, might still score within the normal range on an intelligence test. In addition, two further disadvantages of the use of these tests in a multicenter setting became obvious in this study: First, the effort necessary to administer the test is significant, requiring at least three hours of a trained person's time. Second, the tests cannot be performed in all patients. In our series, 6 of 21 patients could not be carried out, or were incapable of completing the intelligence test. The exclusion of very young children and severely handicapped patients reduced the usefulness of these tests in this particular population. Therefore, our data do not support the use of this test for global assessment of health status or quality of life in a multicenter setting.

The Child Behavior Checklist measures behavior and emotional outcomes and may therefore be used to assess these components of global health status. It requires only moderate effort and is quantitative. The test shares with intelligence testing the problem that not all patients can be assessed, because of language issues. The CBCL is appropriate for children 6 to 18 years of age. Preschool forms are now available for ages 1.5 to 5 years (24); however, these were not available in German at the time this study was conducted. In addition, the ASEBA corporation has recently published an English version of an Adult Self-Report and an Adult Behavior Checklist, appropriate for adults aged from 18 to 59 years; however, this is not available in German (25). The most significant problem raised by our data, however, is the questionable validity of this test as a measure of health status in this particular population. The measure was not designed to assess children with medical illnesses, but rather focuses on assessing clinically significant emotional and behavioral disorders. Therefore, it seems to have a decreased sensitivity for identification of less serious psychological problems that are more characteristic of children with chronic medical illnesses (26). The results of the CBCL did not correlate with any of the other tests (Table III) and the debriefing interviews with the parents confirmed the questionable accuracy. This was most likely caused by neurological symptoms in our patients, which may wrongly suggest psychological, psychiatric, or social problems.

The FMH could be completed in all of our patients; it is quantitative and very easy to perform. The total time required to answer the questions was well below 10 minutes per patient. The questions outlined in Table II are translations of the German items. They are provided here to illustrate this test to the international society, but unfortunately cannot yet be used as an English version of the test, since they have not passed single item language improvements, normalization, or measurements of reliability or validity.

Our study confirmed the previously described correlation of the FMH with other methods (Tables II, III), confirming the convergent validity of the instrument. The greatest disadvantage of the FMH may be its limited sensitivity in detecting specific cognitive defects and slow processing speed. These phenomena were addressed only indirectly in terms of their impact on the independence of the patient in daily life. However, this finding may indicate a useful discriminant validity of the instrument for measurement of health status separate from cognitive abilities.

The ongoing German brain tumor studies include measurements of quality of life in multicentre settings. Based on the data presented here, the FMH has been incorporated into these studies. The FMH is accompanied by two other fast tests for the multicenter data collection: i) finger tapping velocity assessed with a computer based system which has been adapted using the software of a well established rapid eye movement test (27, 28); and ii) the Pediatric Quality of Life Questionnaire [PEQDOL (29)]. The data of the pilot studies for those methods will be published shortly. In a few selected centers, a more extensive battery of tests will be completed by a subset of patients to detect specific cognitive defects.

Acknowledgements
We thank S. Whitsett and H. Ottensmeier for helpful discussions, and Jon E. Meyers for editing the manuscript.

References


Received August 8, 2006
Revised November 9, 2006
Accepted November 16, 2006