



Neurocognitive impairment and patient–proxy agreement on health-related quality of life evaluations in recurrent high-grade glioma patients

Ivan Caramanna^{1,2} · Martin Klein^{1,2}  · Martin van den Bent³ · Ahmed Idbaih⁴ · Wolfgang Wick⁵ · Martin J. B. Taphoorn^{6,7} · Linda Dirven^{6,7} · Andrew Bottomley⁸ · Jaap C. Reijneveld^{2,9,10} on behalf of The EORTC Quality of Life Group and EORTC Brain Tumor Group

Accepted: 5 July 2022 / Published online: 18 August 2022
© The Author(s) 2022

Abstract

Purpose The rate of missing data on patient-reported health-related quality of life (HRQOL) in brain tumor clinical trials is particularly high over time. One solution to this issue is the use of proxy (i.e., partner, relative, informal caregiver) ratings in lieu of patient-reported outcomes (PROs). In this study we investigated patient–proxy agreement on HRQOL outcomes in high-grade glioma (HGG) patients.

Methods Generic and disease-specific HRQOL were assessed using the EORTC QLQ-C30 and QLQ-BN20 in a sample of 501 patient–proxy dyads participating in EORTC trials 26101 and 26091. Patients were classified as impaired or intact, based on their neurocognitive performance. The level of patient–proxy agreement was measured using Lin’s concordance correlation coefficient (CCC) and the Bland–Altman limit of agreement. The Wilcoxon signed-rank test was used to evaluate differences between patients’ and proxies’ HRQOL.

Results Patient–proxy agreement in all HGG patients ($N=501$) ranged from 0.082 to 0.460. Only 18.8% of all patients were neurocognitively intact. Lin’s CCC ranged from 0.088 to 0.455 in cognitively impaired patients and their proxies and from 0.027 to 0.538 in cognitively intact patients and their proxies.

Conclusion While patient–proxy agreement on health-related quality of life outcomes is somewhat higher in cognitively intact patients, agreement in high-grade glioma patients is low in general. In light of these findings, we suggest to cautiously consider the use of proxy’s evaluation in lieu of patient-reported outcomes, regardless of patient’s neurocognitive status.

Keywords Quality of Life · Glioma · Patient–proxy agreement · Neurocognitive functioning · PROs

✉ Martin Klein
m.klein@amsterdamumc.nl

¹ Department of Medical Psychology, Amsterdam UMC, Vrije Universiteit Amsterdam, de Boelelaan 1118, PK 1Y 176, 1081 HZ Amsterdam, The Netherlands

² Brain Tumor Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

³ Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, The Netherlands

⁴ Sorbonne Université, Inserm, CNRS, UMR S 1127, Institut du Cerveau et de la Moelle Épineuse, ICM, AP-HP, Hôpitaux Universitaires La Pitié Salpêtrière - Charles Foix, Service de Neurologie 2-Mazarin, 75013 Paris, France

⁵ Clinical Cooperation Unit Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ) & Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany

⁶ Department of Neurology, Leiden University Medical Centre, Leiden, The Netherlands

⁷ Haaglanden Medical Centre, The Hague, The Netherlands

⁸ Quality of Life Department, European Organization for Research and Treatment of Cancer, Brussels, Belgium

⁹ Department of Neurology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

¹⁰ Department of Neurology, Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, The Netherlands

Plain English summary

Since aggressive brain tumors are associated with a dismal prognosis, it is pivotal to evaluate and monitor health-related quality of life in these patients during the disease course. In addition to neurological deficits, these patients often have impaired neurocognitive functioning, which is associated with missing, or possibly inadequate, health-related quality of life evaluations and consequently necessitates involvement of proxies to obtain such information. With this study we investigated to what extent patients' health-related quality of life reported by proxies reflects the patients' self-reports. We found that proxies' evaluation of the patients' health-related quality of life was weakly associated with patients' self-reports, regardless of them having neurocognitive deficits or not. We therefore suggest using proxy's evaluations of high-grade glioma patients' function with caution.

Introduction

The prognosis for high-grade glioma patients is quite dire, with a median survival time of 15 months and a usually rapid decline in general health. Therefore, it is not surprising that health-related quality of life (HRQOL) has become an important secondary outcome measure in high-grade glioma clinical trials, as a measure of patients' functioning [1]. Traditional clinical trial outcomes such as progression-free survival or overall survival do not provide a complete picture of the patient's functioning and well-being. Therefore, outcomes such as HRQOL and neurocognitive functioning are now typically included in brain tumor clinical trials, to better capture functioning and well-being. Unfortunately, many brain tumor clinical trials still suffer from a substantial amount of missing HRQOL data over time. [2] Restricting analyses only to patients able to offer complete patient-reported outcomes (PROs) might introduce a bias in clinical trials, since important HRQOL evaluations of patients with poor neurological and/or neurocognitive functioning might be missing [3]. A possible solution to this problem has been proposed in resorting to evaluations offered by partners, relatives, or informal caregivers, collectively referred to as "proxies" as substitute data in the analysis of missing scores reported by patients.

However, it is unclear to what extent proxy-reported outcomes are representative of the patients' self-perceived outcomes. Previous studies in low-grade glioma patients showed that the level of neurocognitive functioning determines the degree of patient-proxy agreement. Moderate to

high patient-proxy agreement was found in neurocognitively intact patients, while in patients with neurocognitive impairment, patient and proxy ratings differed regarding emotional functioning [4–7]. In general, there seems to be agreement between patients and proxies concerning physical functioning and symptoms [6], but the same cannot be said regarding less visible issues, such as mood and emotional functioning [8].

Furthermore, the debate concerning the subjectivity of HRQOL is still open. While HRQOL ratings are by definition subjective and should in principle be reported by the patient him- or herself [9], HGG patients often face neurological and neurocognitive deterioration that could force clinicians to resort to proxy ratings because of the inability of the patient to do so.

The aim of this study is to investigate patient-proxy HRQOL agreement in a large sample of patients with recurrent HGG. Previous findings of a study conducted on low-grade glioma patients, in which part of the authors involved in the present study collaborated, suggest neurocognitive impairment as an influencing factor for HRQOL patient-proxy agreement. Therefore, we divided patients participating in two EORTC-coordinated clinical trials in neurocognitively impaired and intact and used an approach similar to the one previously implemented in the study by Ediebah and colleagues. Our pre-trial hypothesis was that we expected neurocognitively impaired patients to have lower levels of patient-proxy agreement than neurocognitively intact patients [4–7].

Patients and methods

The initial sample of EORTC trials 26101 and 26091 combined consisted of 731 patients. Most patients had prior chemotherapy and radiotherapy and, in both trials, evaluation prior to randomization and every 12 weeks thereafter included neurocognitive, HRQOL, and full clinical assessment.

In addition to the criteria set for inclusion in the two clinical trials shortly described underneath [10, 11], only HGG patients (i.e., WHO grade III and grade IV) were selected for this study. Since data have been collected prior to 2016, the 2007 WHO tumor classification was used, selecting only WHO grade III and IV tumors [12]. All variables were measured within two weeks prior to randomization. We determined a time window of ± 7 days between neurocognitive functioning (NCF) evaluation and QLQ-C30 and QLQ-BN20 administration to assure concurrent measurements, given the one-week time frame of the QLQ questions.

EORTC trial 26101 (EudraCT number 2009–017,422–39) was a randomized phase III trial investigating whether

the combination of lomustine plus bevacizumab compared to lomustine alone would result in better overall survival in glioblastoma patients with first progression [10]. EORTC 26091 (EudraCT number 2010–023,218-30) was a randomized, open-label phase II trial comparing temozolomide alone to the combination of temozolomide and bevacizumab in patients with a first recurrence of a locally diagnosed WHO grade II or III glioma without 1p/19q co-deletion [11].

Ethics

These trials were approved by the institutional review boards and ethics committees of all participating centers and the respective authorities. The trials were completed according to the Declaration of Helsinki and all patients provided written informed consent.

Neurocognitive assessment

Neurocognitive functioning (NCF) was assessed using a widely accepted clinical trial battery for testing NCF in patients with intracranial or extracranial tumors and were selected because of their wide use in clinical trials and their sensitivity to the impact of tumor and treatment-related variables [13, 14]. This battery consists of the Hopkins Verbal Learning Test–Revised (HVLTR) [15] for total recall, delayed recall, and delayed recognition, which indexes verbal learning and memory, the Trail Making Test (TMT part A and part B) [16], which measures attention, speed, and mental flexibility; and the Controlled Oral Word Association Test (COWAT) [17] test, which evaluates the spontaneous production of words under restricted search conditions. These tests were administered by centrally trained and certified health-care personnel, e.g., research nurses, and neuropsychologists.

Health-related quality of life assessment for patients

HRQOL was measured using the generic EORTC QLQ-C30 questionnaire [18] and the EORTC QLQ-BN20 module specific for brain tumor patients [19]. The former is a 30-item questionnaire developed to assess the quality of life of cancer patients. The latter is a 20-item module which tackles problems specific to brain tumor, its treatment, and consequences.

The QLQ-C30 is divided in functioning and symptoms scales, while the QLQ-BN20 is a symptom-only questionnaire. In functioning scales, the higher the scores, the better the functioning, and in symptoms scales, the opposite is true, a higher score indicates more of the symptoms.

HRQOL questionnaires were filled out at the hospital when patients had scheduled visits. Patients completed the

questionnaire in the clinic, ideally in a quiet, private room; questionnaires were given to the patient before meeting with the physician, ensuring that the patient had enough time to complete the questionnaire. If the patient received a therapy, the questionnaire was filled out before administration of the treatment. The questionnaire could not be taken home and/or mailed.

Health-related quality of life assessment for proxies

Consenting patients were requested to identify a significant other (i.e., spouse or other person in close relationship to the patient), whom physicians asked to participate in the study. The significant others were also provided with verbal and written information on the study.

Patients' proxies were asked to complete the EORTC QLQ-C30 and EORTC QLQ-BN20 at each assessment point at the same time as the patient, at baseline and at 12-weekly follow-ups. Proxies were also instructed to complete the questionnaire trying to put themselves in the shoes of the patients since the questions were formulated always in first person.

Statistical analysis

Descriptive statistics of the sample were calculated, mean and standard for continuous data, and count and or percentages for binary data. For each of the six NCF outcome measures (i.e., 1) HVLTR Total Recall, (2) HVLTR Delayed Recall, (3) HVLTR Delayed Recognition, (4) TMT Part A, (5) TMT Part B, and (6) COWA), raw scores (RS) were calculated [15–17].

Raw scores of the six NCF test outcomes were transformed into Z-scores using available normative data. [15–17] A deviation of -1.5 SD or more from the Z-score mean was used as cut-off to define NCF impairment. Based on the presence of impaired test outcomes, patients were consecutively divided into two groups. Patients who had no impairment on any of the six test outcomes were defined as 'intact,' while patients showing at least one impaired test were defined 'impaired.'

The QLQ-C30 and QLQ-BN20 are questionnaires based on a Likert scale answer system and multi-item and single-item subscales are formed addressing general functioning as well as symptoms. A higher score on a functioning scale corresponds to better functioning, and a higher score on a symptom scale correspond to more symptoms.

Patients with more than half of the indices of neurocognitive functioning (NCF), EORTC Quality of Life Core Questionnaire (QLQ-C30), or Quality of life brain module (QLQ-BN20) evaluations unavailable were excluded from the analyses.

QLQ-C30 and QLQ-BN20 raw scores were transformed into a linear scale ranging from 0 to 100 [20]. Mean differences and standard deviations between patients and proxy were calculated. The proportion of patient–proxy dyads whose difference was within 0, 10, 20, and more than 20 units was summarized. Then, scores of patients and proxies on all QLQ-C30 and QLQ-BN20 scales were compared using Lin's concordance correlation coefficient (CCC) and the Bland–Altman limit of agreement. The Wilcoxon signed-rank test was used to evaluate differences between patients' and proxies' HRQOL.

The Wilcoxon signed-rank test was used to compare the distributions of the patients and proxies scores looking for differences and more importantly to identify eventual systematic bias. Such bias can be caused for instance by a higher median for proxies scores, compared to patients scores.

The Bland–Altman indicates the range within which 95% of all differences in ratings are expected to fall, assuming distribution normality. It was implemented to compare patient–proxy agreement by offering plausible ranges for differences in scores [21–23].

As last CCC was used to test the concordance between patient and proxy ratings. A score below 0.40 indicated poor to fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and 0.81–1.00, excellent agreement [21].

Results

From the initial cohort of 731 patients and 601 proxies, 127 patients were excluded due to completely or extensively missing NCF, QLQ-C30, and/or QLQ-BN20 evaluations and 103 patients did not meet the histological criteria of WHO grade III and IV high-grade gliomas. After the exclusion of those patients with extensively missing data and their significant others together with them, a total of 501 patients were selected: 470 with first recurrence of glioblastoma (EORTC 26101) and 31 patients with first recurrence of a locally diagnosed WHO grade II or III glioma without 1p/19q co-deletion (EORTC 26091). Patients included for the current analyses ($n = 501$) had a median age of 56 years (21–82) and 202 (37.3%) were female. Further detailed clinical information can be found in Table 1.

Agreement between patients and proxies

Table 2 summarizes the QLQ-C30 AND QLQ-BN20 outcomes for all 501 patients meeting the inclusion criteria, regardless of their neurocognitive status.

Differences in scores of patients and proxies were observed on various functioning and symptoms scales, with

patients reporting a higher score (better functioning/more symptoms) than their proxies on Emotional functioning, Cognitive functioning, Nausea and Vomiting, Dyspnea, and Appetite loss scales of the QLQ-C30 and on Seizures, Itchy skin, and Bladder control scales of the QLQ-BN20.

The opposite held true, with patients reporting lower scores (worse functioning/less symptoms) than their proxies on Fatigue and Insomnia scales of the QLQ-C30 and Future uncertainty, Motor dysfunction, Communication deficit, and Drowsiness scales of the QLQ-BN20.

Lin's CCC showed a poor to fair agreement ranging from $r = 0.082$ (Nausea and vomiting) to $r = 0.46$ (Financial Difficulties).

As last the Bland–Altman limit of agreement revealed low agreement between patient and proxies in all HRQOL domains with few exceptions: Global health, Cognitive functioning, Diarrhea, and for the QLQ-BN20, Seizures.

The difference between patients and proxies was calculated, and the proportion within 0 (perfect agreement), 10, 20, and more than 20 units was summarized with a range of 13.6% future uncertainty to 79.2% (Seizures), 0% (Role functioning, Cognitive functioning etc.) to 25.2% (Emotional functioning), 0% (Insomnia, Appetite loss etc.) to 37.7% (Dyspnea), and 0% (Dyspnea) to 55.5% (Drowsiness), respectively.

HRQOL agreement between neurocognitively impaired patients and proxies

In total 94 patients were neurocognitively intact, while 407 were impaired according to our definition. Differences in scores of neurocognitively impaired patients and their proxies were observed on more than half of the functioning and symptoms scales, with patients reporting a higher score than their proxies on the Physical functioning, Role functioning, Emotional functioning, Cognitive functioning, Nausea and vomiting, and Financial difficulties scales of the QLQ-C30 and on the Seizures and Itchy skin scales of the QLQ-BN20.

On the other hand, neurocognitively impaired patients reported lower scores than their proxies on the Fatigue and Insomnia scales of the QLQ-C30 and on the Future uncertainty, Motor Dysfunction, Communication deficit, and Drowsiness scales of the QLQ-BN20.

Lin's CCC showed poor to fair agreement with the exception of the Financial difficulties score (that showed moderate agreement), ranging from $r = 0.088$ (Nausea and vomiting) to $r = 0.452$ (Financial Difficulties).

Again, the Bland–Altman limit of agreement revealed low agreement between patient and proxies in all HRQOL domains, except for Constipation, Diarrhea on the QLQ-C30, and Seizures on the QLQ-BN20.

The difference between patients and proxies was calculated, and the proportion within 0, 10, 20 and more than 20

Table 1 Baseline clinical characteristics

		N°	%		
Age	Median = 56 (21–82)				
Gender					
	Female	202	37.3		
	Male	271			
WHO performance status					
	0	176	35.1		
	1	271	54.1		
	2	54	10.8		
Histology					
	Glioblastoma	436	87		
	Astrocytoma WHO grade III	30	6		
	Glioblastoma with oligodendroglial component	21	4.2		
	Gliosarcoma	8	1.6		
	Giant cell glioblastoma	3	0.6		
	Missing/unknown	3	0.6		
Hemisphere					
	Left	Bilateral	Right		
Tumor location					
	Total	210	23	232	
	Frontal	60	5	65	26
	Temporal	72	0	79	30.1
	Parietal	16	0	20	7.1
	Occipital	20	1	22	8.6
	Other	19	16	10	9
	Multi-site	23	1	366	12
	Missing	36			7.1
AED (Anti-epileptic drugs)					
	Yes			361	66.5
	No			181	33.3
	Missing			1	0.2
Corticosteroids					
	Yes			250	49.9
	No			251	50.1
Neurocognitive impairment					
	Intact			94	18.76
	Impaired			407	81.24

WHO World Health Organization, AED Anti-epileptic drugs

units was summarized with a range of 15% (Future uncertainty) to 78.1% (Seizures), 0% (Role functioning, Cognitive functioning etc.) to 39.6% (Dyspnea), 0% (Insomnia, Appetite loss, etc.) to 37.7% (Dyspnea), 0% (Dyspnea) to 55.8% (Drowsiness), respectively.

Table 3 shows the scores of neurocognitively impaired patients and their proxies.

HRQOL agreement between neurocognitively intact patients and proxies

As can be seen in Table 4, the analysis of the scores of neurocognitively intact patients and their proxies showed significant differences on five functioning and symptoms scales, with patients reporting a higher score than their proxies on

the Nausea and Vomiting and Dyspnea scales of the QLQ-C30 and on the Itchy skin scale of the QLQ-BN20, while showing lower score (worse functioning/less symptoms) than their proxies on the Physical functioning and Role functioning of the QLQ-C30.

Lin's CCC ranged from poor to moderate with the lowest agreement on the Appetite loss $r=0.027$ and the highest on the Diarrhea scale $r=0.538$.

The Bland–Altman limit of agreement revealed agreement between patient and all functioning scales, Physical functioning, Role functioning, Emotional functioning Cognitive functioning, Social functioning, Fatigue, Pain, and Diarrhea on the QLQ-C30 and Visual, Motor dysfunction, Headache, Seizures, Drowsiness, itchy Skin, Weakness in, and Bladder Control on the QLQ-BN20.

The difference between patients and proxies was calculated, and the proportion within 0, 10, 20, and more than 20 units was summarized with a range of 7.4% (Future uncertainty) to 78.7% (Hair loss), 0% (Role functioning, Cognitive functioning, etc.) to 25.5% (Emotional functioning, Future uncertainty), 0% (Insomnia, Appetite loss, etc.) to 31.9% (Cognitive functioning), and 0% (Dyspnea) to 55.3% (Insomnia), respectively (Table 4).

Discussion

In this study, we aimed at assessing patient–proxy HRQOL agreement in a large sample of high-grade glioma (HGG) patients with and without neurocognitive impairment. To achieve this, we compared the baseline scores of patients and proxies from the EORTC trial 26101 and 26091 on the QLQ-C30 and QLQ-BN20 questionnaires. Our findings primarily suggest that in general there is little agreement between HGG patients and proxies on generic and disease-specific HRQOL outcomes.

These results are only partially in line with other studies on patient–proxy agreement in brain cancer patients. [7, 24] Indeed, when looking at the general agreement of HGG patients with their proxies and comparing it to the results published by Brown and colleagues in a similar population and by Sneeuw and colleagues on a generic cancer population, the agreement reported in our sample is lower. Using a similar statistical approach, the first study reported an ICC between patients and proxies greater than 0.5 on 80% of the measurements; the second showed ICCs ranging from 0.46 to 0.73, indicating a moderate to good level of agreement between patients and proxies. In our study, the agreement ranged from 0.08 to 0.46, with only two scales surpassing the 0.40 threshold that indicates the transition from poor/fair to moderate agreement.

According to the literature comparing patients and proxies evaluations in brain tumor patients [4, 6, 23,

24], neurocognitive impairment is considered to affect patient–proxy agreement. Therefore, we expected neurocognitive impairment to be an influencing factor. However, the findings of the present study do not offer such crystal-clear difference between cognitively impaired and intact patients. While it is true that intact patients showed moderate level of agreement and impaired patients reached a similar level of agreement on only two subscales, the difference between the two groups was not as profound as in other studies. Even though the number of scales showing moderate agreement was higher in neurocognitively intact patients, the lower and upper limit of Lin's CCC range across all scales did not differ much between neurocognitively impaired and intact patients.

Results obtained using the Wilcoxon signed-rank test, as well as those using the Bland–Altman limits of agreement follow a similar pattern, respectively, with less scales showing significant differences and more scales with good levels of agreement for cognitively intact patients than impaired ones, without determining a clear difference.

It is hard to determine whether our expectations were met since neurocognitively impaired patients showed lower levels of patient–proxy agreement than neurocognitively intact patients, but agreement was poor altogether, independently from patients' neurocognitive functioning. The results stress how HRQOL evaluation from patients and proxies are far from aligned and this offers the chance to discuss how this divergence can be interpreted. Perfect patient–proxy agreement is unlikely, and differences in scores, depending on the direction of the difference, determine the interpretation: patients showing higher functioning scores and lower symptoms scores than proxies are considered to underestimate their condition and proxies scoring lower on functioning scales and higher on symptoms scales than patients are considered to overestimate patients' status. The opposite interpretation is possible as well, with proxies underestimating or overestimating patient's functioning and symptoms.

We believe that this way of interpreting the difference in scoring is inadequate, since it implies that one of the two perspectives must be wrong. Clearly, this deviates from the purpose itself of evaluating HRQOL which is a subjective concept by nature. We expected the present study to confirm, as reported in the literature, agreement between patients and proxy, or perhaps smaller differences on those scales concerning aspects of physical functioning and symptoms [6], and discrepancy or greater differences over scales and symptoms related to mood and emotional functioning [8]. The difference in mean scores of patients and proxies observed in this study support this pattern. Indeed, we believe that it could be easier for a proxy to recognize patient's physical distress or functioning impairment in the activities of daily living rather than perceive mood changes or being emotionally empathic.

Table 2 Patient-proxy agreement measured by Lin's CCC and Wilcoxon sign-ranked test for the whole sample

All Patients	Mean Proxy	Mean Patients	Mean Difference (SD)	% Points within the 95% limit of agreement (LL-UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
<i>QLQ-C30</i>											
Global health	61.80	61.70	0.1 (26.05)	96.2 (-50.96 to 51.17)	18.8	23.2	16.2	39.8	2.2	0.255 (0.171 to 0.336)	0.747
Physical functioning	73.92	75.60	-1.68(27.26)	93.6 (-55.1 to 51.75)	17.8	22.8	26.8	32.0	0.8	0.380 (0.303 to 0.453)	0.063
Role functioning	61.60	63.69	-2.09(38.01)	94.4 (-76.59 to 72.41)	25.3	0.0	23.6	49.7	1.4	0.287 (0.204 to 0.366)	0.195
Emotional functioning	63.03	68.97	-5.94(26.25)	93 (-57.39 to 45.51)	17.0	25.2	16.0	39.5	2.4	0.389 (0.313 to 0.460)	0.001
Cognitive functioning	62.25	67.92	-5.67(31.34)	95.6 (-67.09 to 55.75)	27.1	0.0	33.4	37.6	2.0	0.353 (0.275 to 0.436)	0.001
Social functioning	65.47	68.48	-3.01(34.41)	93.6 (-70.45 to 64.44)	28.1	0.0	28.0	41.4	2.6	0.362 (0.283 to 0.608)	0.062
Fatigue	42.97	38.68	4.29(29.13)	93.2 (-52.8 to 61.37)	18.6	0.4	29.4	50.5	1.2	0.318 (0.238 to 0.394)	0.001
Nausea and vomiting	3.49	5.71	-2.22(15.84)	91.2 (-33.27 to 28.83)	69.9	0.0	20.4	8.8	1.0	0.082 (-0.002 to 0.164)	0.001
Pain	16.53	16.19	0.34 (25.66)	93 (-49.96 to 50.63)	47.3	0.0	25.2	26.8	0.8	0.391 (0.314 to 0.462)	0.688
Dyspnea	11.63	14.55	-2.93(27.52)	92.8 (-56.86 to 51.01)	60.1	0.0	37.7	0.0	2.2	0.247 (0.164 to 0.327)	0.016
Insomnia	28.15	23.72	4.43 (34.88)	92.8 (-63.94 to 72.8)	44.7	0.0	0.0	52.9	2.4	0.315 (0.234 to 0.392)	0.007
Appetite loss	8.49	11.38	-2.9 (25.42)	92.4 (-52.72 to 46.93)	67.5	0.0	0.0	31.4	1.2	0.203 (0.118 to 0.284)	0.013
Constipation	11.44	11.65	-0.21 (25.66)	94.2 (-50.51 to 50.09)	62.1	0.0	0.0	32.8	5.2	0.342 (0.260 to 0.419)	0.999
Diarrhea	6.01	6.08	-0.07 (18.64)	97.4 (-36.61 to 36.47)	72.5	0.0	0.0	21.6	6.0	0.257 (0.171 to 0.340)	0.976
Financial difficulties	16.21	18.15	-1.94 (29.2)	90.2 (-59.17 to 55.29)	61.5	0.0	0.0	34.6	4.0	0.460 (0.387 to 0.527)	0.150
<i>QLQ-BN20</i>											
Future uncertainty	42.05	38.67	3.38 (30.20)	94.2 (-55.82 to 62.58)	13.6	20.6	18.0	45.1	2.8	0.320 (0.238 to 0.397)	0.014
Visual disorders	15.34	15.41	-0.07 (22.64)	90.6 (-44.44 to 44.31)	36.3	0.6	27.0	32.8	3.4	0.338 (0.257 to 0.414)	0.617
Motor dysfunction	25.67	21.53	4.13 (27.1)	94 (-48.99 to 57.26)	27.3	0.6	27.8	41.8	2.6	0.454 (0.382 to 0.520)	0.001

Table 2 (continued)

All Patients	Mean Proxy	Mean Patients	Mean Difference (SD)	% Points within the 95% limit of agreement (LL-UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
Communication deficit	28.29	24.54	3.75 (31.47)	91.4 (– 57.94 to 65.43)	27.7	0.0	26.0	43.7	2.6	0.395 (0.318 to 0.467)	0.008
Headache	18.56	19.6	– 1.04 (30.39)	92.2 (– 60.59 to 58.52)	53.7	0.0	0.0	42.8	3.6	0.364 (0.284 to 0.439)	0.438
Seizures	4.18	6.7	– 2.51 (20.66)	96 (– 43.01 to 37.99)	79.2	0.0	0.0	16.2	4.6	0.250 (0.168 to 0.328)	0.010
Drowsiness	31.60	28.42	3.18 (33.35)	92.6 (– 62.19 to 68.56)	40.7	0.0	0.0	55.5	3.8	0.326 (0.244 to 0.403)	0.036
Hair loss	8.9	8.6	0.29 (24.27)	93.6 (– 47.28 to 47.85)	68.9	0.0	0.0	24.0	7.2	0.332 (0.248 to 0.410)	0.806
Itchy skin	8.08	11.99	– 3.91 (27.28)	91.4 (– 57.37 to 49.56)	66.1	0.0	0.0	29.4	4.6	0.232 (0.148 to 0.313)	0.001
Weakness in legs	18.13	16.38	1.75 (32)	87.6 (– 60.96 to 64.47)	55.5	0.0	0.0	39.6	5.0	0.281 (0.197 to 0.361)	0.285
Bladder control	10.3	13.26	– 2.96 (30.57)	89.8 (– 62.88 to 56.97)	63.3	0.0	0.0	33.6	3.2	0.209 (0.123 to 0.292)	0.037

% 0 (cases of no difference between patient and proxy evaluation); % < 10 (cases with less than 10 points of difference between patient and proxy evaluation); % < 20 (cases with less than 20 points of difference between patient and proxy evaluation); % > 20 (cases with more than 20 points of difference between patient and proxy evaluation)

CCC Concordance Correlation Coefficient, LL Lower limit, UL Upper limit, CI Confidence interval

Table 3 Patient-proxy agreement measured by Lin's CCC and Wilcoxon sign-ranked test for neurocognitively impaired patients

Impaired patients	Mean proxy	Mean patients	Mean difference (SD)	% Points within the 95% limit of agreement (LL-UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
<i>QLQ-C30</i>											
Global health	60.01	60.87	- 0.86(25.98)	92.3 (- 51.78 to 50.06)	17.6	24.8	15.7	39.8	2.5	0.251 (0.157 to 0.341)	0.758
Physical functioning	71.25	74.55	- 3.3(27.92)	93.4 (- 58.03 to 51.44)	16.3	22.1	26.5	34.4	0.7	0.377 (0.290 to 0.457)	0.004
Role functioning	58.27	63.13	- 4.86(39.2)	94.1 (- 81.7 to 71.97)	23.1	0	24.1	51.4	1.5	0.248 (0.155 to 0.337)	0.011
Emotional functioning	61.72	68.34	- 6.62(27.23)	93.2 (- 59.99 to 46.75)	15.5	25.1	16.2	41	2.2	0.372 (0.287 to 0.451)	0.001
Cognitive functioning	59.44	66.25	- 6.81(32.6)	92 (- 70.7 to 57.08)	24.1	0	33.7	40.3	2	0.325 (0.237 to 0.408)	0.001
Social functioning	63.19	66.62	- 3.43(35.3)	93.6 (- 72.63 to 65.76)	26.8	0	27.8	43.2	2.2	0.348 (0.259 to 0.431)	0.06
Fatigue	45.14	39.3	5.84(29.87)	92.9 (- 52.7 to 64.39)	17.7	0	29	52.6	0.7	0.289 (0.199 to 0.373)	0.001
Nausea and vomiting	3.58	5.15	- 1.57(14.63)	92.3 (- 30.24 to 27.1)	70	0	20.9	8.4	0.7	0.088 (- 0.008 to 0.182)	0.018
Pain	17.07	16.91	0.12(26.21)	93.2 (- 51.24 to 51.49)	44.7	0	27	27.8	0.5	0.366 (0.280 to 0.447)	0.871
Dyspnea	12.75	14.58	- 1.83(27.86)	93.6 (- 56.45 to 52.78)	58.7	0	39.6	0	1.7	0.248 (0.154 to 0.336)	0.178
Insomnia	28.41	22	6.42(34.37)	94.1 (- 60.95 to 73.78)	45.9	0	0	52.3	1.7	0.311 (0.222 to 0.395)	0.001
Appetite loss	9.12	11.52	- 2.40(25.18)	93.4 (- 51.77 to 46.98)	67.1	0	0	31.7	1.2	0.237 (0.144 to 0.326)	0.076
Constipation	11.48	10.88	0.6(24.32)	95.2 (- 47.05 to 48.26)	63.6	0	0	31.2	5.2	0.352 (0.261 to 0.436)	0.58
Diarrhea	6.49	6.06	0.43(19.61)	97 (- 38.01 to 38.87)	71.5	0	0	23.1	5.4	0.207 (0.110 to 0.300)	0.629
Financial difficulties	16.49	19.21	- 2.72(29.94)	90.2 (- 61.4 to 55.96)	60.9	0	0	35.4	3.7	0.452 (0.370 to 0.527)	0.084
<i>QLQ-BN20</i>											
Future uncertainty	42.94	39.52	3.41(30.38)	94.8 (- 56.13 to 62.96)	15	19.4	17.7	45.5	2.5	0.327 (0.237 to 0.411)	0.032
Visual disorders	16.15	16.49	0.34(23.81)	90 (- 46.33 to 47.01)	32.7	0.7	27.5	35.6	3.4	0.310 (0.218 to 0.396)	0.828
Motor dysfunction	28.38	23.54	4.84(28.15)	92.3 (- 50.33 to 60.01)	23.1	0.7	28	45.5	2.7	0.428 (0.347 to 0.504)	0.001

Table 3 (continued)

Impaired patients	Mean proxy	Mean patients	Mean difference (SD)	% Points within the 95% limit of agreement (LL-UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
Communication deficit	31.88	26.88	5.01(32.56)	91.6 (– 58.81 to 68.82)	25.1	0	25.6	46.7	2.7	0.395 (0.310 to 0.474)	0.003
Headache	18.36	20.49	– 2.13(31.33)	92 (– 63.53 to 59.28)	52.8	0	0	43.5	3.7	0.337 (0.247 to 0.422)	0.166
Seizures	4.57	6.82	– 2.25(20.54)	96.6 (– 42.51 to 38.02)	78.1	0	0	16.7	5.2	0.290 (0.198 to 0.376)	0.034
Drowsiness	33.33	29.92	3.41(34.01)	94.3 (– 63.24 to 70.06)	40.3	0	0	55.8	3.9	0.305 (0.213 to 0.392)	0.047
Hair loss	9.15	9.06	0.09(25.16)	93.6 (– 49.22 to 49.39)	66.6	0	0	25.6	7.9	0.311 (0.216 to 0.399)	0.944
Itchy skin	8.33	10.99	– 2.66(26.5)	93.2 (– 54.60 to 49.27)	66.6	0	0	28.7	4.7	0.231 (0.135 to 0.322)	0.035
Weakness in legs	19.69	17.53	2.16(33.78)	86.8 (– 64.05 to 68.37)	52.3	0	0	42.5	5.2	0.250 (0.155 to 0.340)	0.283
Bladder control	11.36	14.24	– 2.88(31.36)	90.4 (– 64.36 to 58.59)	61.2	0	0	35.4	3.4	0.225 (0.129 to 0.316)	0.067

% 0 (cases of no difference between patient and proxy evaluation); % < 10 (cases with less than 10 points of difference between patient and proxy evaluation); % < 20 (cases with less than 20 points of difference between patient and proxy evaluation); % > 20 (cases with more than 20 points of difference between patient and proxy evaluation)

CCC Concordance Correlation Coefficient, LL Lower limit, UL Upper limit, CI Confidence interval

Table 4 Patient–proxy agreement measured by Lin’s CCC and Wilcoxon sign-ranked test for neurocognitively intact patients

Intact patients	Mean proxy	Mean patients	Mean difference (SD)	% Points within the 95% limit of agreement (LL–UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
<i>QLQ-C30</i>											
Global health	69.44	65.32	4.21 (26.11)	92.3 (– 46.97 to 55.39)	25.5	16	18.1	29.4	1.1	0.218 (0.027 to 0.394)	0.148
Physical functioning	85.53	80.17	5.36 (22.99)	96.2 (– 39.7 to 50.43)	24.5	24.5	27.7	21.3	1.1	0.305 (0.120 to 0.470)	0.032
Role functioning	75.99	66.13	9.86 (9.72)	98.1 (– 48.39 to 68.1)	35.1	0	23.1	42.6	1.1	0.448 (0.281 to 0.588)	0.004
Emotional functioning	68.77	71.73	– 2.96 (21.32)	95.2 (– 44.76 to 38.83)	23.4	25.5	14.9	33	3.2	0.468 (0.293 to 0.613)	0.132
Cognitive functioning	74.46	75.18	– 0.72 (24.7)	97.1 (– 49.13 to 47.68)	40.4	0	31.9	25.5	2.1	0.412 (0.228 to 0.568)	0.881
Social functioning	75.55	76.66	– 1.11 (30.27)	96.2 (– 60.44 to 58.22)	34	0	28.7	33	4.3	0.356 (0.163 to 0.523)	0.771
Fatigue	33.33	35.95	– 2.63 (24.53)	96.2 (– 50.7 to 45.45)	22.3	2.1	30.9	41.5	3.2	0.437 (0.260 to 0.585)	0.364
Nausea and vomiting	3.08	8.15	– 5.07 (20.19)	90.4 (– 44.64 to 34.49)	69.1	0	18.1	10.6	2.1	0.074 (– 0.087 to 0.231)	0.012
Pain	14.31	13.04	1.27 (23.21)	95.2 (– 44.22 to 46.76)	58.5	0	17	22.3	2.1	0.490 (0.320 to 0.630)	0.535
Dyspnea	6.67	14.44	– 7.78 (25.5)	92.3 (– 57.76 to 42.21)	66	0	29.8	0	4.3	0.244 (0.072 to 0.402)	0.005
Insomnia	26.96	31.46	– 4.49 (35.95)	90.4 (– 74.96 to 65.98)	39.4	0	0	55.3	5.3	0.342 (0.151 to 0.508)	0.246
Appetite loss	5.73	10.75	– 5.02 (26.44)	91.3 (– 56.84 to 46.81)	69.1	0	0	29.8	1.1	0.027 (– 0.165 to 0.217)	0.053
Constipation	11.23	14.98	– 3.75 (30.75)	92.3 (– 64.01 to 56.52)	55.3	0	0	39.4	5.3	0.316 (0.123 to 0.485)	0.332
Diarrhea	3.88	6.2	– 2.33 (13.32)	100 (– 28.44 to 23.79)	76.6	0	0	14.9	8.5	0.538 (0.374 to 0.669)	0.109
Financial difficulties	14.98	13.48	1.5 (25.58)	94.2 (– 48.64 to 51.63)	63.8	0	0	30.9	5.3	0.498 (0.327 to 0.638)	0.653
<i>QLQ BN- 20</i>											
Future uncertainty	38.15	34.9	3.24 (29.56)	94.2 (– 54.69 to 61.17)	7.4	25.5	19.1	43.6	4.3	0.262 (0.061 to 0.442)	0.187
Visual disorders	10.37	12.21	– 1.83 (16.67)	97.1 (– 34.5 to 30.84)	25.1	0	24.5	20.2	3.2	0.477 (0.304 to 0.620)	0.413
Motor dysfunction	14.01	12.92	1.09 (21.92)	96.2 (– 41.87 to 44.05)	45.2	0	26.6	25.5	2.1	0.474 (0.301 to 0.617)	0.517

Table 4 (continued)

Intact patients	Mean proxy	Mean patients	Mean difference (SD)	% Points within the 95% limit of agreement (LL-UL)	% Within 0 points	% With < 10	% With < 20	% With > 20	Missing	CCC (95% CI)	Wilcoxon <i>p</i> values
Communication deficit	12.80	14.49	- 1.69 (25.73)	94.2 (- 52.12 to 48.73)	39.4	0	27.7	30.9	2.1	0.068 (- 0.134 to 0.263)	0.868
Headache	19.41	15.75	3.66 (25.56)	96.2 (- 46.43 to 53.75)	57.4	0	0	39.4	3.2	0.500 (0.335 to 0.635)	0.188
Seizures	2.53	6.15	- 3.62 (21.23)	95.2 (- 45.23 to 37.99)	84	0	0	13.8	2.1	0.036 (- 0.130 to 0.200)	0.114
Drowsiness	24.17	21.97	2.2 (30.55)	96.2 (- 57.68 to 62.08)	42.6	0	0	54.3	3.2	0.372 (0.182 to 0.535)	0.469
Hair loss	7.77	6.66	1.11 (20.26)	92.3 (- 46.97 to 55.39)	78.7	0	0	17	4.3	0.437 (0.257 to 0.587)	0.664
Itchy skin	7.03	16.29	- 9.25 (29.99)	96.2 (- 39.7 to 50.43)	63.8	0	0	31.9	4.3	0.243 (0.067 to 0.405)	0.007
Weakness in legs	11.48	11.48	0.00 (22.89)	98.1 (- 48.39 to 68.1)	69.1	0	0	26.6	4.3	0.437 (0.255 to 0.590)	0.873
Bladder control	5.79	9.05	- 3.26 (27.09)	95.2 (- 44.76 to 38.83)	72.3	0	0	25.5	2.1	0.051 (- 0.144 to 0.243)	0.275

% 0 (cases of no difference between patient and proxy evaluation); % < 10 (cases with less than 10 points of difference between patient and proxy evaluation); % > 20 (cases with more than 20 points of difference between patient and proxy evaluation)

CCC Concordance Correlation Coefficient, LL Lower limit, UL Upper limit, CI Confidence interval

Altogether the lack of patient–proxy agreement over HRQOL reflected in the results of the present study, and the importance of not considering a patient untrustworthy solely due to its condition stress the lack of a tool to establish PROs reliability.

It is important to mention that there are several limitations to this study. The fact that treatment course and disease duration prior to inclusion may have been different between patients in the two trials might have impacted sample homogeneity. Additionally, a selection bias due to missing NCF data might have affected the results.

Data concerning the level of kinship of proxies was not recorded at the EORTC headquarters. At the time of the design of these studies, it was regarded to be counterproductive to register demographic data on the proxies, as that would have required informed consent by the proxy as well, possibly negatively influencing recruitment rates of these EORTC studies. No information about specific procedures used to assess tests which were independently completed was recorded. However, in each institution, one person was appointed as the responsible for the local organization of HRQOL data collection. This could have been a physician, data manager, (research) nurse, or a psychologist.

The percentage of mean differences between dyads (0, 10, 20, or more than 20) might have been influenced by the number of possible scores on a scale. [23] Furthermore, no direct measure of mood, which has been shown to offer even more insight into patient–proxy levels of agreement, was collected [8]. Generalizability might be limited due to the selection bias which characterizes clinical trial populations in general. Finally and importantly, our definition of impairment is arbitrary. It is possible that by grading the extent of neurocognitive impairment in levels rather than in a dichotomic variable, results might be different, unfortunately in our case this was not possible. Nevertheless, a sensitivity analysis raising the threshold for neurocognitive dysfunction per test (> 2 SD) was performed. By exacerbating the definition of neurocognitive impairment from what is commonly considered the impairment threshold in the clinical environment, we hoped to include only those with an impaired performance even if on only one of the NCF tests. Results show how raising the neurocognitive impairment threshold produced little difference compared to the methodology implemented in the present study and suggests that the threshold adopted in the present study does not limit its message.

The aim of this study was to assess patient–proxy HRQOL agreement in a large sample of high-grade glioma (HGG) patients with and without neurocognitive impairment. The intrinsic subjectivity of health-related quality of life evaluation makes it difficult to establish what the ‘truth’ is. Our initial assumption was based on a syllogism for which a cognitive intact patient could be considered a reliable source of his/her own quality of life and a caregiver should be a reliable observer, at least for those scales describing functioning aspects and observable symptoms.

However, the question that would follow is a predictable one: Would it be legitimate not to rely on patients evaluation of his/her own well-being due to neurocognitive impairment?

The results of this study suggest that the level of patient–proxy agreement in HGG patients is low in general. When patients were divided into cognitively impaired and intact, these latter showed agreement with their proxies on more scales of the questionnaires, but the level of agreement remained low, suggesting, in contrast with previous literature that cognitive impairment might influence but not preclude agreement. We hope that future studies will tackle the lack of a quantitative measure of reliability of PROs in patients at risk for neurocognitive impairment. Moreover, in light of these findings, we would suggest to cautiously consider the use of proxy’s evaluation in lieu of PROs at least until a measure to establish reliability is developed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03197-w>.

Acknowledgements The authors are grateful to the medical and nursing staff of all the facilities which took part in the clinical trials, the patients, and their significant others for their cooperation in this study.

Funding The study was made possible by unconditional Grant 007/2016 of the EORTC Quality of Life Group.

Declarations

Conflict of interest No authors have a potential conflict of interest.

Research involving human and animals rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Taphoorn, M. J., Stupp, R., Coens, C., Osoba, D., Kortmann, R., van den Bent, M. J., Mason, W., Mirimanoff, R. O., Baumert, B. G., Eisenhauer, E., Forsyth, P., Bottomley, A., European Organisation for Research and Treatment of Cancer Brain Tumour Group; EORTC Radiotherapy Group; National Cancer Institute of Canada

- Clinical Trials Group. (2005). Health-related quality of life in patients with glioblastoma: A randomised controlled trial. *Lancet Oncology*, 6(12), 937–944. [https://doi.org/10.1016/S1470-2045\(05\)70432-0](https://doi.org/10.1016/S1470-2045(05)70432-0)
2. Meyers, C. A., & Hess, K. R. (2003). Multifaceted end points in brain tumor clinical trials: Cognitive deterioration precedes MRI progression. *Neuro-Oncology*, 5(2), 89–95. <https://doi.org/10.1093/neuonc/5.2.89>
 3. Dirven, L., Reijneveld, J. C., Aaronson, N. K., Bottomley, A., Uitdehaag, B. M. J., & Taphoorn, M. J. B. (2013). Health-related quality of life in patients with brain tumors: Limitations and additional outcome measures. *Current Neurology and Neuroscience Reports*. <https://doi.org/10.1007/s11910-013-0359-y>
 4. Ediebah, D. E., Reijneveld, J. C., Taphoorn, M. J., Coens, C., Zikos, E., Aaronson, N. K., Heimans, J. J., Bottomley, A., Klein, M., EORTC Quality of Life Department and Patient Reported Outcome and Behavioral Evidence (PROBE). (2017). Impact of neurocognitive deficits on patient-proxy agreement regarding health-related quality of life in low-grade glioma patients. *Quality of Life Research*, 26(4), 869–880. <https://doi.org/10.1007/s11136-016-1426-z>
 5. Gehring, K., Taphoorn, M. J. B., Sitskoorn, M. M., & Aaronson, N. K. (2015). Predictors of subjective versus objective cognitive functioning in patients with stable grades II and III glioma. *Neuro-Oncology Practice*, 2(1), 20–31. <https://doi.org/10.1093/nop/npu035>
 6. Armstrong, T. S., Wefel, J. S., Gning, I., Acquaye, A., Vera-Bolanos, E., Gilbert, M. R., Cleeland, C. S., & Mendoza, T. (2012). Congruence of primary brain tumor patient and caregiver symptom report. *Cancer*, 118(20), 5026–5037. <https://doi.org/10.1002/cncr.27483>
 7. Brown, P. D., Decker, P. A., Rummans, T. A., Clark, M. M., Frost, M. H., Ballman, K. V., Arusell, R. M., & Buckner, J. C. (2008). A prospective study of quality of life in adults with newly diagnosed high-grade gliomas: Comparison of patient and caregiver ratings of quality of life. *American Journal of Clinical Oncology: Cancer Clinical Trials*, 31(2), 163–168. <https://doi.org/10.1097/COC.0b013e318149f1d3>
 8. Rooney, A. G., McNamara, S., MacKinnon, M., Fraser, M., Rampling, R., Carson, A., & Grant, R. (2013). Screening for major depressive disorder in adults with glioma using the PHQ-9: A comparison of patient versus proxy reports. *Journal of Neuro-Oncology*, 113(1), 49–55. <https://doi.org/10.1007/s11060-013-1088-4>
 9. Osoba, D. (1994). Lessons learned from measuring health-related quality of life in oncology. *Journal of Clinical Oncology. American Society of Clinical Oncology*. <https://doi.org/10.1200/JCO.1994.12.3.608>
 10. Wick, W., Gorlia, T., Bendszus, M., Taphoorn, M., Sahm, F., Harting, I., Brandes, A. A., Taal, W., Domont, J., Idhah, A., Campone, M., Clement, P. M., Stupp, R., Fabbro, M., Le Rhun, E., Dubois, F., Weller, M., von Deimling, A., Golfopoulos, V., ... van den Bent, M. J. (2017). Lomustine and bevacizumab in progressive glioblastoma. *New England Journal of Medicine*, 377(20), 1954–1963. <https://doi.org/10.1056/NEJMoA1707358>
 11. van den Bent, M. J., Klein, M., Smits, M., Reijneveld, J. C., French, P. J., Clement, P., de Vos, F. Y. F., Wick, A., Mulholland, P. J., Taphoorn, M. J. B., Lewis, J., Weller, M., Chinot, O. L., Kros, J. M., de Heer, I., Verschuere, T., Coens, C., Golfopoulos, V., Gorlia, T., & Idhah, A. (2018). Bevacizumab and temozolomide in patients with first recurrence of WHO grade II and III glioma, without 1p19q co-deletion (TAVAREC): a randomised controlled phase 2 EORTC trial. *The Lancet Oncology*, 19(9), 1170–1179. [https://doi.org/10.1016/S1470-2045\(18\)30362-0](https://doi.org/10.1016/S1470-2045(18)30362-0)
 12. Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvett, A., Scheithauer, B. W., & Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica*. Springer. <https://doi.org/10.1007/s00401-007-0243-4>
 13. Wefel, J. S., Cloughesy, T., Zazzali, J. L., Zheng, M., Prados, M., Wen, P. Y., Mikkelsen, T., Schiff, D., Abrey, L. E., Yung, W. K., Paleologos, N., Nicholas, M. K., Jensen, R., Vredenburgh, J., Das, A., & Friedman, H. S. (2011). Neurocognitive function in patients with recurrent glioblastoma treated with bevacizumab. *Neuro-oncology*, 13(6), 660–668. <https://doi.org/10.1093/neuonc/nor024>
 14. Wefel, J. S., Saleeba, A. K., Buzdar, A. U., & Meyers, C. A. (2010). Acute and late onset cognitive dysfunction associated with chemotherapy in women with breast cancer. *Cancer*, 116(14), 3348–3356. <https://doi.org/10.1002/cncr.25098>
 15. Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins verbal learning test - Revised: Normative data and analysis of inter-form and test-retest reliability. *Clinical Neuropsychologist*, 12(1), 43–55. <https://doi.org/10.1076/clin.12.1.43.1726>
 16. Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
 17. Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). *Benton Controlled Oral Word Association Test: Reliability and Updated Norms*. *Archives of Clinical Neuropsychology* (Vol. 1). Retrieved from <https://academic.oup.com/acn/article-abstract/11/4/329/1614>
 18. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., Haes, J. C. J. M. D., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). The European organization for research and treatment of cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85(5), 365–376. <https://doi.org/10.1093/jnci/85.5.365>
 19. Osoba, D., Aaronson, N. K., Muller, M., Sneeuw, K., Hsu, M. A., Yung, W. K., Brada, M., & Newlands, E. (1996). The development and psychometric validation of a brain cancer quality-of-life questionnaire for use in combination with general cancer-specific questionnaires. *Quality of Life Research*, 5(1), 139–150. <https://doi.org/10.1007/BF00435979>
 20. Fayers, P., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). *EORTC QLQ-C30 Scoring Manual*. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer
 21. Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
 22. Bland, J. M., & Altman, D. G. (n.d.). *STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT*.
 23. Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>
 24. Sneeuw, K. C., Aaronson, N. K., Sprangers, M. A., Detmar, S. B., Wever, L. D., & Schornagel, J. H. (1997). Value of caregiver ratings in evaluating the quality of life of patients with cancer. *Journal of Clinical Oncology*, 15(3), 1206–1217. <https://doi.org/10.1200/JCO.1997.15.3.1206>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.