European Journal of Cancer 212 (2024) 114335

Contents lists available at ScienceDirect

European Journal of Cancer

journal homepage: www.ejcancer.com

Instrumental activities of daily living in neuro-oncology: International validation of the EORTC IADL-BN32 questionnaire

Q. Oort [a,b], J.C. Reijneveld [a,b,c], S.A.M. Sikkes [d,e], J.A.F. Koekkoek [f,g], F. Boele [h,i], T. Young [j], C. Brannan [j], T. Chalk [j], A. Talacchi [k], A. Mazzotta [k], Y. Narita [l], H. Sato [l,m], Y. Miyakita [l,n], O. Shamieh [o,p,q], W. Alrjoob [q], A. Pace [r], D. Petranovic [s], M. Ploh [s], A. Capela [t,u], J. Silva [t,u], M.J. Hjermstad [v,w], T. Urbanic Purkart [x], C. Seidel [y], N. Talhi [y], J. Pichler [z], I. Höllmüller [z], L. Brown [aa], M. Hand [aa], M. Klein [a,ab], N.K. Aaronson [ac], B.M.J. Uitdehaag [b], M.J.B. Taphoorn [f,g,*], L. Dirven [f,g], on behalf of the EORTC Quality of Life Group and the EORTC Brain Tumor Group

[a] *Cancer Center Amsterdam, Brain Tumor Center, Amsterdam, the Netherlands*
[b] *Department of Neurology & Brain Tumor Center Amsterdam, Amsterdam University Medical Centers (Location VUmc), Amsterdam, the Netherlands*
[c] *Department of Neurology, Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, the Netherlands*
[d] *Alzheimer Center, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands*
[e] *Vrije Universiteit Amsterdam, Faculty of Behavioural and Movement Sciences (FGB), Department of Clinical Developmental & Clinical Neuropsychology, Amsterdam, the Netherlands*
[f] *Department of Neurology, Haaglanden Medical Center, The Hague, the Netherlands*
[g] *Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands*
[h] *Leeds Institute of Medical Research, St James's University Hospital, Leeds LS9 7TF, United Kingdom*
[i] *Leeds Institute of Health Sciences, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, United Kingdom*
[j] *East & North Hertfordshire NHS Trust Incorporating Mount Vernon Cancer Centre, Northwood, United Kingdom*
[k] *Department of Neurosurgery, Azienda Ospedaliera San Giovanni Addolorata, Roma, Italy*
[l] *Department of Neurosurgery and Neuro-oncology, National Cancer Center Hospital, Tokyo, Japan*
[m] *Department of Nursing, Teikyo Heisei University, Tokyo, Japan*
[n] *Cancer Institute Hospital of JFCR, Tokyo, Japan*
[o] *Department of Palliative Medicine, King Hussein Cancer Center (KHCC), Amman, Jordan*
[p] *School of Medicine, University of Jordan, Amman, Jordan*
[q] *Center for Palliative & Cancer Care in Conflict (CPCCC), KHCC, Amman, Jordan*
[r] *Neuro-Oncology Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy*
[s] *Department of Hematology, Clinical Hospital Center Rijeka, Rijeka, Croatia*
[t] *Associação de Investigação de Cuidados de Suporte em Oncologia (AICSO), Portugal*
[u] *Centro Hospitalar Vila Nova de Gaia, Espinho, Portugal*
[v] *European Palliative Care Research Centre (PRC), Department of Oncology, Oslo University Hospital and Institute of Clinical Medicine of University of Oslo, Oslo, Norway*
[w] *Regional Advisory Unit for Palliative Care, Department of Oncology, Oslo University Hospital, Oslo, Norway*
[x] *Department of Neurology and Neuroradiology, Vascular and Interventional Radiology, Medical University of Graz, Austria*
[y] *Department of Radiation Oncology, University of Leipzig Medical Center, Leipzig, Germany*
[z] *Institut für Innere Medizin mit Neuroonkologie, Kepler Universitätsklinikum GmbH, Linz, Austria*
[aa] *John Eastwood Hospice, Sutton-in-Ashfield, United Kingdom*
[ab] *Department of Medical Psychology, Amsterdam University Medical Centers (Location VUmc), Amsterdam, the Netherlands*
[ac] *Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands*

ARTICLE INFO

Keywords:
Instrumental activities of daily living
IADL
Daily functioning

ABSTRACT

Background: Neurocognitive impairments are common in patients with a brain tumour, and may negatively impact on functioning in daily life, particularly on *instrumental* activities of daily living (IADL). The EORTC IADL-BN32 questionnaire was developed to measure IADL in this patient population.

* Correspondence to: Leiden University Medical Center, Department of Neurology, PO BOX 9600, 2300 RC Leiden, the Netherlands.
E-mail address: m.j.b.taphoorn@lumc.nl (M.J.B. Taphoorn).

https://doi.org/10.1016/j.ejca.2024.114335
Received 12 August 2024; Received in revised form 13 September 2024; Accepted 14 September 2024
Available online 21 September 2024

Brain tumour
Glioma
Brain metastasis

Methods: In this international validation study, we evaluated the EORTC IADL-BN32 questionnaire on several psychometric properties in a large sample of patients with a primary or metastatic brain tumour. We administered the 32-item questionnaire three times: at 'baseline', after 2 weeks and after 3 months. Procedures were in accordance with EORTC Quality of Life Group module development guidelines.
Results: In total, 326 patients participated in the study. A bifactor scale structure showed satisfactory model fit measures, with five multi-item scales and two single items, and an IADL sum score. The internal consistency of the multi-item scales ranged from good to excellent (range Cronbach's α: 0.86–0.97). We found significant differences in scale scores between patients with and without neurocognitive impairments or complaints, supporting the construct validity. Initial cross-cultural validity analyses showed indications of item response biases for certain items. Analyses indicated moderate to good test-retest agreement (intraclass correlation coefficient > 0.70) between baseline and the 2-week follow-up assessment for all but one scale. Deterioration of EORTC IADL-BN32 scale scores were consistent with clinically relevant deterioration on other functional measures with small to large effect sizes, however, subgroup sample sizes were small.
Conclusion: Overall, the EORTC IADL-BN32 questionnaire exhibited adequate to excellent psychometric properties. Cross-cultural validity and responsiveness should be further explored.

1. Introduction

Brain tumours comprise a heterogeneous group of tumours. There are two main types of brain tumours: primary brain tumours, which originate in the brain, and secondary, or metastatic brain tumours, which arise from other cancer sites and metastasize to the brain [1,2]. Patients with malignant primary and metastatic brain tumours have less than favourable prognoses, from several years for patients with low-grade gliomas to merely several months for patients with metastatic brain tumours [3]. Moreover, the overall symptom burden in patients with a brain tumour is significant. Impairments in physical and neurocognitive abilities can be severe and can be present during the entire disease trajectory [4–7]. These impairments can greatly impede a patient's ability to function in their everyday life. Indeed, studies in other patient populations with neurocognitive impairments, such as those suffering from dementia [8], suggest that deterioration of neurocognitive functioning is associated with more problems with, particularly, instrumental activities of daily living (IADL), such as household activities or using a computer.

Although patients with brain tumours may be prone to problems with everyday activities due to neurocognitive deficits, no reliable and valid brain-tumour specific IADL questionnaire is available as of yet. Given the generally progressive and incurable nature of brain tumours, it is imperative to assess, monitor, and preserve as long as possible. An important first step toward the development of an IADL measure for brain tumour patients involved the evaluation of the applicability of a validated IADL questionnaire developed for dementia patients (i.e., Amsterdam IADL Questionnaire© (A-IADL-Q)) [8–11] in the brain tumour population [12]. Results indicated the need of an IADL questionnaire specifically for patients with brain tumours, as they reported difficulties in daily functioning not covered by the A-IADL-Q. Therefore, it was decided to develop and validate an IADL questionnaire specifically for patients with a brain tumour following the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG) module development guidelines [13]. Questionnaire development consists of four phases: I) generation of relevant issues, II) operationalization of the issues into a set of items, III) pre-testing the questionnaire; and IV) larger scale, international field testing of the psychometric properties of the questionnaire. Phases I–III have been completed [14] and resulted in a preliminary EORTC IADL-BN32 questionnaire. In this phase IV study we aimed to evaluate the psychometric properties of the EORTC IADL-BN32 questionnaire in a large international sample of patients with brain tumours.

2. Methods

2.1. Study design and patient population

This international, multicentre validation study recruited adult patients (≥ 18 years) with a brain tumour who visited a Neuro-Oncology, Radiation Oncology or Medical Oncology in- or outpatient clinic. Patients were eligible to participate if they had a histologically confirmed low- or high-grade glioma according to the WHO 2016 classification criteria, or a metastatic brain tumour and a histologically confirmed primary tumour. In addition, patients were expected to remain clinically stable in at least the first two weeks after recruitment, and had to have a life expectancy of at least 3 months. The aim was to recruit a well-balanced group of patients with respect to the target population. Therefore, we focused on recruitment of a similar number of patients with respect to *tumour type* (low-grade glioma [WHO grade 2], high-grade glioma [WHO grade 3–4], or brain metastases). In addition, as we hypothesized that neurocognitive deficits would impact the performance of IADL, we also aimed to recruit a similar number of patients with and without neurocognitive impairments (based either on the impression of the primary health care professional (HCP) or neurocognitive assessments [if available]). Lastly, all participants were required to have sufficient understanding of the main language of the country in which they live, enabling completion of questionnaires and interviews. In addition, if available, a proxy (i.e., defined as persons in close contact with the patient, such as the partner, spouse, child or parent of the patient was recruited. Proxy data is not reported in this paper, as psychometric analyses were based on patient data only, except for the acceptability of the questionnaire (see Supplementary file 1 - Proxies). All participants were informed of the study procedures and signed an informed consent prior to participation.

In accordance with EORTC QLG module development guidelines [13], patients were recruited from three main European geographical regions (Northern Europe [Austria, Germany, Norway and The Netherlands]; Southern Europe [Italy and Portugal]; Eastern Europe [Croatia]), an English-speaking country [United Kingdom], and non-European countries [Japan and Jordan]) to ensure cross-cultural applicability.

2.2. Assessment schedule and instruments

Patients received questionnaires three times: at 'baseline' (i.e. first assessment for a patient, but this may be at any moment in the disease trajectory), 2 weeks (2-week follow-up (FU)) and 3 months (3-month FU) later. At baseline, patients received four questionnaires:

- A study-specific *General information form* for sociodemographic information (e.g. patient's gender, age, level of education);
- The phase III-version of the *EORTC IADL-BN32 questionnaire* [14]. The EORTC IADL-BN32 consists of 32 items comprising five multi-item scales and two single-item scales (see also Fig. 1). Items were scored on a 4–point Likert scale, ranging from 'not at all' to 'very much'. Although the phase III-version of the questionnaire also included the option 'not applicable' for each item, this was omitted

in this phase IV validation study due to analytical issues in phase III. Scoring of the multi-item and single-item scales followed the EORTC scoring manual [15], with a linear transformation of scores into a score ranging from 0 to 100 with a higher score indicating more problems with activities. The EORTC IADL-BN32 questionnaire can be requested through the website of the EORTC Quality of Life Group: https://qol.eortc.org/questionnaires/;

- The *MOS Cognitive Functioning Scale-Revised* (MOS CFS-R) is a subscale of the Medical Outcome Study (MOS) Core Measures of Health-related quality of life questionnaire [16] and used to assess subjective neurocognitive complaints. The subscale consists of 6 items with a 6-point Likert scale. Total scores range from 6 to 36 points. The Medical Outcome Study (MOS) Core Measures of Health-related quality of life questionnaire can be requested through the website of QualityMetric: https://www.qualitymetric.com/health-surveys/sf-cognitive-functioning-scale/;
- A study-specific *debriefing questionnaire* (5 items, see Supplementary file 2 – Debriefing questionnaire) to determine the feasibility and acceptability of the questionnaire.

For the second and third assessment, patients only completed the EORTC IADL-BN32 and the MOS CFS-R. Treating physicians provided relevant tumour- and treatment-related data from patients' medical records (e.g. the presence of 'intracranial progression' based on tumour growth assessed with imaging). In phase 3 of the development of the questionnaire it became clear that administration of a neurocognitive test battery to assess the patients' neurocognitive status was difficult, and therefore we choose other sources of information in this phase IV validation study. Neurocognitive status was determined by the treating physicians, a neuropsychologist or, if available, based on a neuropsychological assessment if performed as part of patient care. Furthermore, the treating physicians evaluated the patient's performance level by completing the *Karnofsky Performance Status* (KPS) scale [17] and the

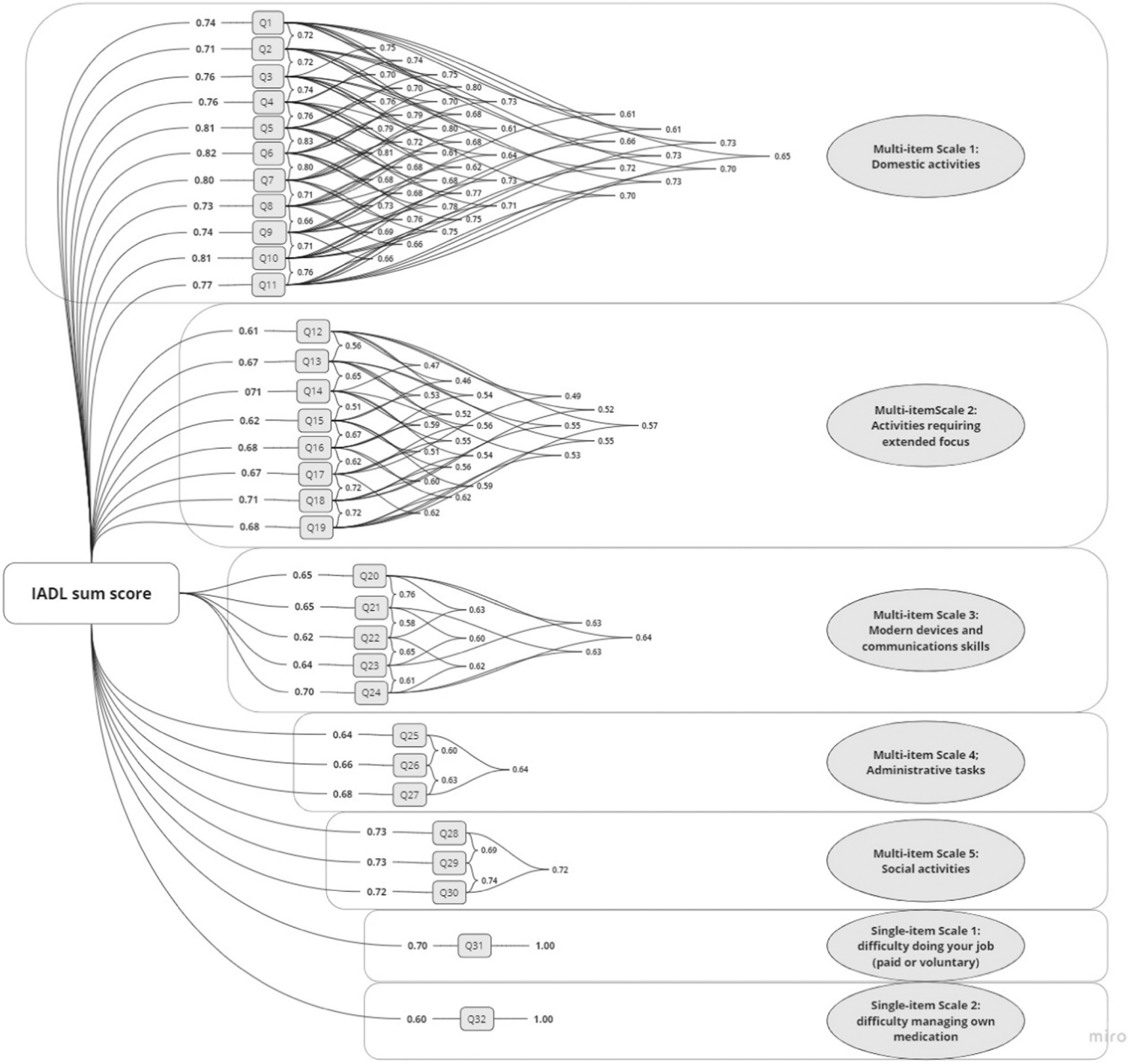

Fig. 1. EORTC IADL BN-32 bi-factor scale structure with inter-item correlations.

patient's level of dependence by completing the *Barthel Index* (BI) [18] to assess problems with basic activities of daily living. These outcomes were collected at baseline and at the 3-month FU assessment.

2.3. Statistical analyses

Baseline clinical and sociodemographic characteristics were analysed using descriptive statistics.

2.3.1. Compliance and acceptability

We evaluated compliance rate (i.e., percentage of patients completing all items of the EORTC IADL-BN32) and dropouts at each assessment. Based on the debriefing questionnaire completed by both patients and proxies, items rated as '*confusing or difficult*' or '*upsetting*' by >5 % of participants of the total sample or >5 % of participants from a single geographical region, were more closely reviewed and addressed by either rephrasing the item or, if not possible, excluding the item. Furthermore, we analysed responses from the open comment section qualitatively; if more than five participants indicated a similar issue, we considered revision or exclusion of the item.

2.3.2. Inter-item correlations

Based on the baseline patient data only, we determined inter-item Spearman rank correlations to establish item divergence or item redundancy. Items should not have very weak (<0.2) or very strong correlation (>0.9) with all other items, as this indicates that the item does not measure the same construct or measures the exact same construct, respectively.

2.3.3. Structural validity

We determined the structural validity, defined as the degree to which the scores of the items in a questionnaire are an adequate reflection of the dimensionality of the construct to be measured, by means of a bifactor confirmatory factor analysis (CFA) with baseline patient data. The aim was to confirm the scale structure as identified with the exploratory factor analysis in phase III (i.e., five multi-item scales and two single-items) [14], together with an IADL sum score. The comparative fit index (CFI) and Tucker–Lewis index (TLI) were calculated as measures of incremental model fit indexes [19]. We considered the scale structure to have a satisfactory model fit if the indexes were >0.90 [20]. In addition, a root mean square of approximation (RMSEA) of <0.10 was considered indicative of a reasonable fit [20] and a standardized root mean square residual (SRMR), a measure of the mean absolute correlation residual, of <0.06 was considered a good fit [21].

2.3.4. Internal consistency

We determined the internal consistency of the multi-item scales, defined as the interrelatedness among the items in a multi-item scale, with Cronbach's alpha based on the patient baseline data. Cronbach's alpha <0.5 was considered '*unacceptable*', between 0.5 and 0.59 as '*poor*', between 0.60 and 0.69 as '*questionable*', between 0.70 and 0.79 as '*acceptable*', between 0.80 and 0.89 as '*good*', and ≥ 0.90 as '*excellent*' [19]. We considered a Cronbach's alpha of at least 0.70 as requirement for sufficient internal consistency of a scale [22].

2.3.5. Cross-cultural validity

To gain insight into the level of cross-cultural validity of the EORTC IADL-BN32 questionnaire, we analysed potential item biases in the patient baseline data of the five geographical regions (i.e. the Northern European region [Austria, Germany, Norway and The Netherlands], the Southern European region [Italy and Portugal], Eastern Europe [Croatia], the English-speaking region [United Kingdom] and the non-European region, [Japan and Jordan]) by means of differential item functioning (DIF) analyses with both item response theory (IRT) and non-IRT methods (i.e., the Mantel-Haenszel method [23]), and two IRT methods (i.e., the logistic regression method [24] and Lord's chi-square test method [25]). First, we dichotomized the item data into '*no problems*' (score of 1 on an item) and '*problems*' (i.e. scores 2–4 on an item) to remove the severity aspect and simplify the data. As no 'reference group' could be determined, we compared each geographic region to the other four regions by dichotomizing. As no previous DIF analyses have been conducted on this questionnaire, it is unknown if there are any DIF-free items that could be used as anchor items in the analyses, therefore none were set. We performed descriptive analyses to assess the direction of the discrepancy between regions. If all three DIF methods detect the presence of an item bias for a specific item for a specific geographical region, this would suggest that the item might be measuring different abilities for this geographic region.

2.3.6. Construct validity

The construct validity is defined as the degree to which the scales in a questionnaire are consistent with hypotheses based on the assumption that the questionnaire validly measures the construct to be measured. Since there is no 'gold standard' to measure IADL in patients with brain tumours, the criterion validity could not be assessed and instead construct validation was performed by means of known-groups comparisons. Therefore, we constructed a priori defined hypotheses (Table 1) to demonstrate the discriminatory ability of the questionnaire between groups based on a sociodemographic characteristic (age) and relevant clinical subgroups (tumour type, tumour recurrence/progression before baseline, performance status (KPS), basic ADL [BADL] as assessed with the Barthel Index (BI), neurocognitive status and the level of subjective complaints as assessed with the MOS-CF scale), which were presumed to (significantly) differ between the defined subgroups based on previous studies [26] or expert opinion. Moreover, hypotheses were also constructed with respect to the change in performance of IADL over time between subgroups, for which patients' baseline and the 3-month FU assessments (i.e., to determine a change score) were used (Table 1). We performed Mann Whitney U tests to determine significant differences (in change) between groups. The magnitude of the effect size between two groups was determined by calculating Cohen's *d*. Cohen suggested that an effect size between 0.2 and 0.5 should be considered 'small', between 0.5 and 0.8 'medium' and >0.8 'large' [27].

2.3.7. Test-retest reliability

The test-retest reliability is defined as the degree to which the results are consistent over time. This psychometric property was based on the patients' baseline and the 2-week FU assessments. We analysed relative reliability using intra-class correlation coefficients (ICC; Two-way mixed effect, Absolute Agreement). An ICC of at least 0.70 was considered '*sufficient*' test-retest reliability [22], between 0.75 and 0.90 was defined as '*good*' agreement and between 0.90–1.00 as '*excellent*' agreement [31]. In addition, we determined absolute reliability by means of the standard error of measurement (SEM) and the smallest detectable change for individual subjects (SDC). The lower the SEM and SDC, the more reliable the measure. The SEM and SDC were also expressed as a percentage of the range of possible scores for each measure. This would allow for comparisons of measurement error among different measures by correcting for the different units of the scales. Furthermore, we constructed Bland-Atman plots with limits of agreement (LoA) to depict the group level congruence between the baseline and the 2-week FU assessments.

2.3.8. Responsiveness

We analysed responsiveness to assess the ability of the questionnaire to detect clinically relevant changes over time. Patients' baseline and the 3-month FU assessments were used. To determine whether changes in IADL performance over time were clinically relevant, we used relevant changes in other functional measures as anchor, namely changes in the Karnofsky Performance Status (KPS) score, neurocognitive severity status, subjective neurocognitive complaints, or BI score. We expected that IADL outcomes would show changes in accordance with the

Table 1
Hypotheses for the known-group comparisons.

Sociodemographic	Groups	At baseline	Differences between groups over time (i.e. difference between groups in change scores between baseline and the 3-month follow-up assessment)
Age	Patients 70 and older vs. those younger than 70 years.	Older patients will report having more difficulty performing IADL compared to younger patients.	Older patients were presumed to report significantly more problems in IADL performance over time, compared to younger patients.
Clinical			
Tumour type	Patients with a low-grade glioma vs. patients with a high-grade glioma	Patients with a high-grade glioma will have more difficulty performing IADL than patients with a low-grade glioma.	Patients with faster growing tumour types (i.e. high-grade glioma and brain metastases patients) were presumed to report significantly more problems in IADL performance over time, compared to patients with a low-grade glioma.
	Patients with a low-grade glioma vs. patients with brain metastases	Patients with brain metastases will have more difficulty performing IADL than patients with a low-grade glioma.	
Recurrent or new tumour growth before baseline/ intracranial progression	Recurrent or new tumour growth before baseline vs. no recurrent or new tumour growth before baseline	Patients with recurrent or new tumour growth before baseline will have more difficulty performing IADL than those without recurrent or new tumour growth.	Patients with recurrent or new tumour growth were presumed to report significantly more problems in IADL performance over time, compared to patients without recurrent or new tumour growth.
Performance status (Karnofsky performance status [KPS])	Lower levels of performance status (KPS < 70) vs. higher levels of performance status (KPS ≥ 70)	Patients with KPS < 70 will have more difficulty performing IADL than patients with a KPS ≥ 70.	Patients with a ≥ 10 points decrease[a] in KPS score over time were presumed to report significantly more problems in IADL performance over time, compared to patients without a ≥ 10 points[a] decrease in KPS score. Patients with a ≥ 10 points increase[a] in KPS score over time were presumed to report significantly less problems in IADL performance over time, compared to patients without a ≥ 10 points increase in KPS score.
BADL (Barthel Index [BI])	Independent [BI $= 100$] vs. dependent [BI < 100]	Patients who are dependent will have more difficulty performing IADL than patients who are considered independent.	Patients with a ≥ 10 points decrease[a] in BI score over time were presumed to report significantly more problems in IADL performance over time, compared to without a ≥ 10 points decrease in BI score. Patients with a ≥ 10 points increase[a] in BI score over time were presumed to report significantly less problems in IADL performance over time, compared to patients without a ≥ 10 points increase in BI score.
Neurocognitive status	Neurocognitively impaired vs. neurocognitively unimpaired (based on impression healthcare professional and/ or tests)	Patients who are considered to have neurocognitive impairments will have more difficulty performing IADL than patients who are considered neurocognitively unimpaired.	Patients who were considered to have declined in their neurocognitive functioning were presumed to report significantly more problems in IADL performance over time, compared to patients who were assessed to have not declined (further) in their neurocognitive functioning. Patients who were assessed to have improved in their

(*continued on next page*)

Table 1 (*continued*)

Sociodemographic	Groups	At baseline	Differences between groups over time (i.e. difference between groups in change scores between baseline and the 3-month follow-up assessment)
			neurocognitive functioning were presumed to report significantly less problems in IADL performance over time, compared to patients who were assessed to have not improved (further) in their neurocognitive functioning.
Subjective neurocognitive complaints (MOS Cognitive Functioning Scale-Revised [MOS CFS-R])	High levels of neurocognitive complaints vs. low level of neurocognitive complaints, based on the median MOS CFS-R score of 30.	Patients with a high level of subjective neurocognitive complaints (MOS CFS-R $\leq$ 30) will have more difficulty performing IADL than patients with a low level of neurocognitive complaints (MOS CFS-R $>$ 30).	Patients with a $\geq$ 6 points decrease[a] in MOS CFS-R score over time were presumed to report significantly more problems in IADL performance over time, compared to patients without a $\geq$ 6 points decrease in MOS CFS-R score. Patients with a $\geq$ 6 points increase[a] in MOS CFS-R score over time were presumed to report significantly less problems in IADL performance over time, compared to patients without a $\geq$ 6 points increase in MOS CFS-R score.

[a] For the KPS and BI, an increase or decrease of $\geq$ 10 points on a 0–100 point scale over time was considered clinically meaningful [28,29]. For the MOS CFS-R, an increase or decrease of $\geq$ 6 points on a 6–36 point scale over time seemed clinically meaningful [30].

changes on these functional measures, e.g., patients who showed decline on these functional measures would also decline on the IADL outcomes.

We examined responsiveness using the Wilcoxon signed rank test to measure if the repeated-measure differences were significant. In addition, we determined the standardized mean difference (SMD) and standardized response mean (SRM) between the baseline and 3-month FU assessment to assess the effect sizes of the responsiveness. The SMD effect size is calculated with the standard deviation of the baseline, and the SRM effect size is calculated with the standard deviation of the difference. In general, these effect sizes are also interpreted as Cohen [27] suggested.

All analyses were performed with SPSS version 28.0 (Armonk, NY: IBM Corp). In the above mentioned analyses, no corrections were conducted for multiple testing as we consider the analyses to determine the psychometric properties as exploratory analyses.

3. Results

Baseline sociodemographic and clinical characteristics of patients are presented in Table 2 (more detailed table in Supplementary file 3 – Extensive baseline sociodemographic and clinical characteristics). We recruited 335 patients, of which 326 completed a baseline assessment. As shown in Table 2, the number of patients with different tumour types and with and without neurocognitive problems was relatively well-balanced. However, patients from the Northern European region were overrepresented (38 % of population).

There were statistically significant differences between baseline sociodemographic and clinical variables between regions (see Table 2). Notably, patients from the Eastern European region appeared to have less favourable functional outcomes compared to other regions. The BI scores were significantly more often below 100 (71 %) for Eastern European patients compared to the other regions (range 19–43 %), they were more often assessed as having neurocognitive impairments (79 %) compared to the other regions (30–48 %), and they reported more subjective neurocognitive complaints (mean score 24.6), compared to other regions (range mean scores: 27.1–29.5). The reason that patients from the Eastern European region appeared to have less favourable functional outcomes compared to other regions is most likely due to the higher percentage of patients with either a high-grade glioma (46 %) or metastatic brain tumour (50 %; in total 96 % of the patient population), compared to other regions (range 26–59 % and 4–47 %, respectively and in total a range of 53–72 % of the patient population). However, post-hoc analyses indicated that these differences were not statistically significant.

3.1. Psychometric properties

3.1.1. Compliance and acceptability

Compliance was *adequate*, 83 % (271/326) of patients completed all items of the questionnaire and 95 % (311/326) completed $\geq$ 30 items. For only 11 % (37/326) of patients, one or more scales scores could not be calculated, mostly the single-item scales (92 %; 34/37), but the IADL sum score could be calculated for all patients. For the full study population, patients' dropout rate was 15 % (n = 50) for the 2-week FU assessment and 27 % (n = 88) for the 3-month FU assessment (for details see, Supplementary file 4 – Drop-out rates), but the compliance rate of those who fully completed questionnaire at the 2-week and 3-month FU assessment remained the same, namely 83 % and 84 % respectively. Some items were more often omitted, but by no more than 2 % of the patient population, and some items were more often described as 'Not applicable', including item 11 (*Difficulty taking care of family members (including children)*, n = 14 (4 %)), item 28 (*Difficulty organizing a social activity (*e.g. *a dinner)*, n = 11 (3 %)) and item 31 (*Difficulty doing your job (paid or voluntary)*, n = 21 (6 %)) (see Supplementary file 5 – Compliance rate for more details). None of the items had to be rephrased or removed based on the criteria regarding items being rated as '*confusing or difficult*' or '*upsetting*' (data not shown). The debriefing questionnaire indicated that the average completion time of the EORTC IADL-BN32 was 11.8 min (SD = 13.3 min; ranging from 1.5 to 120 min; missing n = 5). A total of 24 % (n = 78) of the patients needed help filling in the questionnaire. Patients with a high grade glioma or metastatic brain tumour required more frequently help (50 % and 35 %, respectively) compared to patients with a low grade glioma (15 %). Patients who required help were also more often considered

Table 2
Baseline sociodemographic and clinical characteristics of patients.

	All patients
PATIENT CHARACTERISTICS	
Number of patients, N (%)	326
Northern Europe	124 (38 %)
Southern Europe	61 (19 %)
Eastern Europe	24 (7 %)
English-speaking region	27 (8 %)
Non-European region	90 (28 %)
Sex (male), N (%) [Missing]	174 (53 %) [n = 1]
Northern Europe	67 (55 %)
Southern Europe	36 (60 %) [n = 1]
Eastern Europe	7 (29 %)
English-speaking region	17 (61 %)
Non-European region	47 (52 %)
Age (yrs), M (SD) [Missing]	54.4 (13.6) [n = 1]
Northern Europe	56.3 (13.0)
Southern Europe	52.2 (13.4) [n = 1]
Eastern Europe	62.8 (14.3)
English-speaking region	53.6 (13.0)
Non-European region	51.2 (13.7)
Level of Education [1–8][a] **(Low** [1–4]**) [Missing], N (%)**	150 (46 %) [n = 3]
Northern Europe	64 (52 %) [n = 1]
Southern Europe	41 (68 %) [n = 1]
Eastern Europe	11 (48 %) [n = 1]
English-speaking region	13 (46 %)
Non-European region	21 (23 %)
KPS, Median [quartiles 25–75 %]	90 [70–90]
Northern Europe	90 [80–90]
Southern Europe	90 [80–100]
Eastern Europe	75 [60–80]
English-speaking region	70 [60–90]
Non-European region	80 [70–90]
Barthel Index [0–100], Median [quartiles 25–75 %]	100 [95–100]
Northern Europe	100 [100]
Southern Europe	100 [95–100]
Eastern Europe	83 [60–100]
English-speaking region	100 [100]
Non-European region	100 [90–100]
Neurocognitive status (impaired), N (%)	124 (38 %)
Northern Europe	46 (37 %)
Southern Europe	18 (30 %)
Eastern Europe	19 (79 %)
English-speaking region	13 (48 %)
Non-European region	28 (31 %)
Subject neurocognitive complaints (MOS CFS-R) [0–36], Median [quartiles 25–75 %] [Missing]	30 [23–34] [n = 14]
Northern Europe	28 [22–32] [n = 8]
Southern Europe	31 [27–35] [n = 2]
Eastern Europe	24 [18–33] [n = 1]
English-speaking region	31 [22–34] [n = 3]
Non-European region	32 [24–36]
Dominant hand (right), N (%) [Missing]	288 (88 %) [n = 11]
Northern Europe	108 (92 %) [n = 7]
Southern Europe	54 (90 %) [n = 1]
Eastern Europe	20 (91 %) [n = 2]
English-speaking region	22 (85 %) [n = 1]
Non-European region	84 (93 %)
TUMOUR CHARACTERISTICS	
Tumour type, N (%)	
Low-grade glioma	100 (31 %)
Northern Europe	35 (35 %)
Southern Europe	29 (29 %)
Eastern Europe	1 (1 %)
English-speaking region	10 (10 %)
Non-European region	25 (25 %)
Low-grade glioma	
Diffuse astrocytoma	
– *IDH-mutant*	35 (16 %)
– *IDH-wildtype*	7 (3 %)
– *NOS*	9 (4 %)
Oligodendroglioma	
– *IDH-mutant and 1p/19q-codeleted*	40 (18 %)
– *NOS*	5 (2 %)
Oligoastrocytoma, NOS	2 (1 %)
Cerebral ependymoma	1 (0.5 %)
Unable to determine	1 (0.5 %)
High-grade glioma	120 (37 %)
Northern Europe	48 (40 %)
Southern Europe	22 (18 %)
Eastern Europe	11 (9 %)
English-speaking region	16 (13 %)
Non-European region	23 (19 %)
High-grade glioma	
Anaplastic astrocytoma	
– *IDH-mutant*	8 (4 %)
– *IDH-wildtype*	7 (3 %)
– *NOS*	3 (1 %)
Anaplastic oligodendroglioma	
– *IDH-mutant and 1p/19q-codeleted*	6 (3 %)
– *NOS*	1 (0.5 %)
– *Glioblastoma*	
– *IDH-mutant*	9 (4 %)
– *IDH-wildtype*	81 (37 %)
NOS	4 (2 %)
Diffuse midline glioma, H3 K27M-mutant	1 (0.5 %)
Brain metastases	106 (33 %)
Northern Europe	41 (39 %)
Southern Europe	10 (9 %)
Eastern Europe	12 (11 %)
English-speaking region	1 (1 %)
Non-European region	42 (40 %)
Brain metastases	
Number of metastases, Median [range] [Missing]	2 [1–30] [n = 4]
Primary tumour location, N (%)	
Lung	47 (44 %)
Breast	26 (25 %)
Melanoma	12 (11 %)
Other	21 (20 %)
Brain tumour location, N (%) [Missing]	[n = 1]
Frontal	115 (35 %)
Temporal	54 (17 %)
Parietal	27 (8 %)
Occipital	5 (2 %)
Multiple	92 (28 %)
Other	32 (10 %)
Time since diagnosis (months), M (SD) [range] [Missing]	37 (58.5) [0 −378] [n = 1]
Current tumour treatment, N (%)	181 (56 %)
Chemotherapy only (extracranial + intracranial chemotherapy)	102 (56 %)
Radiotherapy only (extracranial + intracranial radiotherapy)	15 (8 %)
Other (either TTF, Proton RT, targeted-, immune-, hormonal therapy)	18 (10 %)
Multiple treatments	46 (25 %)
Previous tumour treatment (multiple options possible), N (%)	
None	21 (6 %)
Biopsy	44 (13 %)
Resection	224 (69 %)
Re-resection	42 (13 %)
Chemotherapy	167 (51 %)
Radiotherapy	216 (66 %)
Other (TTF, Proton RT, targeted-, immune-, hormonal therapy)	37 (11 %)

Abbr. N = Number, M = Mean, SD = Standard Deviation, YRS = Years, KPS = Karnofsky Performance Status, NOS = Not Otherwise Specified, IDH = Isocitrate dehydrogenase, TTF = Tumour-treating fields, RT = Radiotherapy.

[a] The level of education is based on The International Standard Classification of Education (ISCED). Scores range between 1 and 8, with higher score representing a higher level of education. Scores 1–4 are classified as a low level of education.

neurocognitively impaired (72 % vs. 27 %) and were more often dependent in their functioning than patients who did not receive help (i.e. BI < 100; 63 % vs. 37 %).

3.1.2. Inter-item correlations

The inter-item correlations ranged from 0.31–0.83. None of the items had a very weak correlation ($\rho < 0.2$) or very strong correlation

($\rho > 0.9$) with any of the other IADL items (Fig. 1), indicating that questionnaire items represented the same construct, but not to a problematically low or high extent.

3.1.3. Structural validity

The bi-factor CFA model (Fig. 1) had a CFI and TLI of 0.92 and 0.90, respectively, indicating the bi-factor model has satisfactory model fit measures (i.e., both > 0.90). The RMSEA and the SRMR were 0.08 and 0.05, respectively, also indicative of a good fit (i.e., < 0.10 and < 0.06, respectively). These results indicate that the preliminary scale structure developed in phases I-III *with* an IADL sum score has a satisfactory model fit. We used this scale structure in all subsequent analyses.

3.1.4. Internal consistency

The five multi-item scales were evaluated for their internal consistency, which was found to be good to excellent (i.e., Cronbach's $\alpha \geq 0.70$), with Cronbach's α of 0.97, 0.92, 0.90, 0.86 and 0.90 for scales 1–5, respectively.

3.1.5. Cross-cultural validity

We performed differential item functioning (DIF) analyses to detect item response biases for certain geographical regions. Due to the limited number and unequal distribution of patients among the different geographical regions, three broader geographical regions were constructed as described in the Methods section to perform the DIF analyses. See Supplementary file 6 – Differential Item Functioning (DIF) for the item response bias detection per analytical method for each item and per geographical region. Overall, we detected response biases more often for 'Northern Europe' as compared to 'Southern European' and 'Non-European regions'. More specifically, patients from 'Northern Europe' reported on average less often problems with IADL for items 2, 4, 8 and 10 (mean difference (MD) of − 0.06, − 0.06, − 0.09 and − 0.06, respectively), compared to the other regions, and more often problems for four other items (MD = 0.21 for item 16, MD = 0.20 for item 17, MD = 0.20 for item 18 and MD = 0.17 for item 24). Patients from the Non-European region reported less often problems with items 18 and 24 (MD = − 0.23 and MD = − 0.22, respectively), compared to the other regions.

3.1.6. Construct validity

The discriminatory ability of the questionnaire was determined based on known-group comparisons of relevant subgroups. Only the significant results are described in Table 3.

Overall, the EORTC IADL-BN32 questionnaire is reasonably good at discriminating between relevant clinical subgroups, specifically regarding neurocognition subgroups, but also variables related to functioning in daily life (performance status and BADL), showing significant differences with small to large effect sizes (range 0.43–1.45) at baseline. Particularly, the EORTC IADL-BN32 questionnaire is reasonably good at discriminating subgroups based on *deterioration* of neurocognition and functioning in daily life (small to large effect sizes: range 0.42–1.55), however less convincingly so for *improvements* of neurocognitive status and functioning in daily life (small to medium effect sizes: range 0.01–0.79). Improvements in subjective neurocognitive complaints showed medium to large significant effects (range 0.53–1.04), except for Single-item scale 1 (*Difficulty doing your job (paid or voluntary)* with an effect size of 0.11.

In general, we found almost no differences based on age groups, and if so only small effect sizes. Regarding tumour type, results were also less hypothesis affirming, with few significant differences and predominantly small effect sizes (range 0.06–0.60). Moreover, unexpectedly, patients with brain metastases did not report more problems on all or most IADL scales at baseline, compared to patients with LGG, indicating that patients with brain metastases have fewer issues than expected. Patients with tumour recurrence before baseline did show significantly more problems in almost all multi-item IADL scales and the IADL sum score, but the effect sizes were small (range 0.01–0.29). Patients with intracranial tumour progression between the baseline and 3-month FU assessments reported significantly more problems with some IADL scales with medium effect sizes.

3.1.7. Test-retest reliability

We analysed the test-retest reliability to determine the consistency and reproducibility of the IADL outcomes. We assessed the consistency between patient's responses at baseline and at 2-week FU (Table 4), a period in which we expected that the disease remained stable and therefore the level of IADL performance would be similar. The results show that all multi- and single-item scales and the IADL sum score showed moderate to good agreement (ICC range: 0.70–0.80), however Single-item scale 2 (*Difficulty managing own medication)* was not sufficiently reliable (ICC = 0.57). The SEM was between 1 % and 2 % and the SDC ranged between 2 % and 5 % of the total range of the scale score (range 0–100) for all the scales, which is indicative of good reliability. This is also supported by the finding that the differences in responses between the baseline and 2-week FU assessments were comparable (see Supplementary file 7 - Bland-Atman plots).

3.1.8. Responsiveness

No overall significant changes in IADL outcome scores were found between the baseline and 3-month FU assessments and effect sizes were negligible (< 0.20) for both SMD and SRM. There were, however, significant differences over time in IADL performance in relevant clinical subgroups of patients. Results from the responsiveness analyses are shown in Table 5.

The EORTC IADL-BN32 questionnaire showed varying responsiveness based on clinically relevant changes in other functional measures. Although the subgroup sample sizes were small, there are indications that the EORTC IADL-BN32 questionnaire is particularly responsive in detecting deterioration in patient's performance status, BADL, and neurocognitive functioning both assessed by a HCP (neurocognitive status) and the patients themselves (neurocognitive complaints): deterioration in these functional measures reflected poorer performance in most IADL outcomes (although not all statistically significant), with small to large effect sizes. We did not observe a similar level of responsiveness for detecting improvements in IADL performance for the functional measures. Only patients with a decrease in subjective neurocognitive complaints showed an improvement in performance of almost all multi-item IADL scales and the IADL sum score, with small to large effects.

4. Discussion

This international multicentre phase IV validation study has shown that the EORTC IADL-BN32 questionnaire has, overall, adequate to excellent psychometric properties. The acceptability of the questionnaire appeared to be good, as no items had to be rephrased or removed. In terms of feasibility, 24 % of the patients needed help filling in the questionnaire. These patients were mostly patients with more malignant tumour types (HGG and brain metastases), and those with neurocognitive impairments and problems with functioning in daily life. This is not unexpected considering the patient population and should be taken into account when administrating questionnaires in clinical trials and practice.

The internal consistency of the items was within the norms (i.e., inter-item correlations between 0.2 and 0.9) and was considered good to excellent for the multi-item scales. The structural validity of the EORTC IADL-BN32 questionnaire with a bifactor CFA model, reflecting the phase III hypothesized scale structure [14] together with an IADL sum score, did have satisfactory model fit measures. The initial results related to cross-cultural validity showed there were indications of item response biases for some items for certain geographical regions, however, it remains unclear how meaningful these findings are. While item responses biases may be caused by actual regional differences, we cannot rule out

Table 3
Known-group comparisons per IADL scale as well as the IADL sum score of the a priori defined relevant sociodemographic and clinical subgroups. Results of the discriminating ability of the EORTC IADL-BN32 questionnaire scales and IADL sum score between the relevant subgroups are presented for the outcomes at baseline, and for changes over time (both deterioration and improvement) in outcomes.

	Baseline				*Over time*			
Sociodemographic	*N*	*MD*	*d*	*p*	*N*	*MD*	*d*	*p*
Age								
< 70 yo vs. ≥ 70 yo								
Scale 1 (Domestic activities)	274 vs. 49	− 1.23	− 0.04	-	202 vs. 33	6.57	0.27	-
Scale 2 (Activities requiring extended focus)	274 vs. 50	− 4.99	− 0.20	-	202 vs. 35	5.83	0.29	0.04
Scale 3 (Modern devices and communication skills)	273 vs. 50	− 11.08	− 0.45	0.02	202 vs. 34	5.31	0.27	-
Scale 4 (Administrative tasks)	274 vs. 50	− 7.81	− 0.30	-	202 vs. 32	− 2.07	− 0.09	-
Scale 5 (Social activities)	270 vs. 49	− 5.07	− 0.16	-	199 vs. 32	7.16	0.26	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	253 vs. 38	10.31	0.27	-	179 vs. 26	8.91	0.25	-
Single-item scale 2 (Difficulty managing own medication)	273 vs. 49	− 3.28	− 0.10	-	200 vs. 35	6.21	0.22	-
IADL sum score	275 vs. 50	− 4.42	− 0.18	-	202 vs. 35	6.09	0.32	-
Clinical								
Tumour type								
LGG vs. HGG								
Scale 1 (Domestic activities)	100 vs. 118	− 8.31	− 0.31	0.02	85 vs. 85	10.01	0.46	0.02
Scale 2 (Activities requiring extended focus)	100 vs. 119	− 1.57	− 0.06	-	85 vs. 86	5.86	0.32	-
Scale 3 (Modern devices and communication skills)	100 vs. 119	− 9.73	− 0.39	< 0.05	85 vs. 85	8.43	0.50	< 0.01
Scale 4 (Administrative tasks)	100 vs. 119	− 5.00	− 0.19	-	85 vs. 85	7.25	0.35	-
Scale 5 (Social activities)	99 vs. 118	− 3.16	− 0.11	-	85 vs. 84	12.91	0.51	0.01
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	95 vs. 103	− 10.77	− 0.28	-	75 vs. 73	8.68	0.25	0.04
Single-item scale 2 (Difficulty managing own medication)	99 vs. 119	− 19.03	− 0.60	< 0.001	84 vs. 85	1.19	0.05	-
IADL sum score	100 vs. 120	− 6.97	− 0.30	-	85 vs. 86	8.47	0.50	< 0.01
LGG vs. BM								
Scale 1 (Domestic activities)	100 vs. 106	− 11.53	− 0.36	-	85 vs. 66	11.47	0.47	-
Scale 2 (Activities requiring extended focus)	100 vs. 106	4.26	0.18	-	85 vs. 66	2.91	0.15	-
Scale 3 (Modern devices and communication skills)	100 vs. 105	− 2.03	− 0.09	-	85 vs. 67	4.67	0.26	-
Scale 4 (Administrative tasks)	100 vs. 106	1.85	0.08	0.03	85 vs. 65	0.96	0.05	-
Scale 5 (Social activities)	99 vs. 103	1.61	0.05	-	85 vs. 63	10.99	0.45	0.02
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	95 vs. 94	− 2.07	− 0.06	-	75 vs. 58	6.90	0.20	-
Single-item scale 2 (Difficulty managing own medication)	99 vs. 105	− 7.68	− 0.28	-	84 vs. 67	2.78	0.11	-
IADL sum score	100 vs. 106	− 3.21	− 0.14	-	85 vs. 67	7.23	0.40	0.03
Tumour recurrence/intracranial progression								
No recurrence/progression vs. recurrence/progression								
Scale 1 (Domestic activities)	207 vs. 114	− 3.79	− 0.12	-	210 vs. 24	19.09	0.79	< 0.01
Scale 2 (Activities requiring extended focus)	207 vs. 115	− 7.03	− 0.29	< 0.01	211 vs. 25	3.80	0.19	-
Scale 3 (Modern devices and communication skills)	206 vs. 115	− 5.67	− 0.23	< 0.01	210 vs. 25	10.19	0.53	-
Scale 4 (Administrative tasks)	207 vs. 115	− 5.95	− 0.23	0.01	209 vs. 24	8.20	0.36	-
Scale 5 (Social activities)	204 vs. 113	− 6.76	− 0.22	< 0.01	206 vs. 24	3.63	0.13	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	183 vs. 106	− 4.93	− 0.13	-	184 vs. 20	18.59	0.53	< 0.05
Single-item scale 2 (Difficulty managing own medication)	206 vs. 114	− 0.47	− 0.01	-	210 vs. 24	4.17	0.15	-
IADL sum score	208 vs. 115	− 5.40	− 0.22	< 0.01	211 vs. 25	11.16	0.59	0.03
Performance status (KPS score)								
KPS < 70 vs. KPS ≥ 70								
Scale 1 (Domestic activities)	45 vs. 279	40.76	− 1.45	< 0.001				
Scale 2 (Activities requiring extended focus)	45 vs. 280	22.26	− 0.94	< 0.001				
Scale 3 (Modern devices and communication skills)	44 vs. 280	28.26	− 1.22	< 0.001				
Scale 4 (Administrative tasks)	45 vs. 280	22.53	− 0.91	< 0.001				
Scale 5 (Social activities)	44 vs. 276	25.83	− 0.87	< 0.001				
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	39 vs. 253	33.48	− 0.91	< 0.001				
Single-item scale 2 (Difficulty managing own medication)	45 vs. 278	29.72	− 0.97	< 0.001				
IADL sum score	46 vs. 280	30.98	− 1.42	< 0.001				
ΔKPS ≥ − 10 points vs. < − 10 points								
Scale 1 (Domestic activities)					46 vs. 187	− 17.02	− 0.73	< 0.01
Scale 2 (Activities requiring extended focus)					46 vs. 189	− 12.47	− 0.65	0.02
Scale 3 (Modern devices and communication skills)					46 vs. 188	− 9.50	− 0.50	0.04
Scale 4 (Administrative tasks)					45 vs. 187	− 12.53	− 0.58	0.02
Scale 5 (Social activities)					43 vs. 186	− 16.16	− 0.62	< 0.01
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					37 vs. 167	− 14.72	− 0.42	-
Single-item scale 2 (Difficulty managing own medication)					45 vs. 188	− 18.75	− 0.70	< 0.001
IADL sum score					46 vs. 189	− 14.46	− 0.80	< 0.001
ΔKPS ≥ + 10 points vs. < + 10 points								
Scale 1 (Domestic activities)					38 vs. 195	0.26	0.01	-
Scale 2 (Activities requiring extended focus)					38 vs. 197	4.27	0.22	-
Scale 3 (Modern devices and communication skills)					38 vs. 196	6.92	0.36	0.03
Scale 4 (Administrative tasks)					38 vs. 194	6.34	0.29	-
Scale 5 (Social activities)					38 vs. 191	8.40	0.32	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					35 vs. 169	2.67	0.08	-
Single-item scale 2 (Difficulty managing own medication)					38 vs. 195	6.80	0.24	-
IADL sum score					38 vs. 197	4.09	0.22	-
BADL (BI score)								
BI < 100 vs. BI = 100								
Scale 1 (Domestic activities)	100 vs. 224	36.14	1.35	< 0.001				

(*continued on next page*)

Table 3 (*continued*)

Sociodemographic	*Baseline*				*Over time*			
	N	*MD*	*d*	*p*	*N*	*MD*	*d*	*p*
Scale 2 (Activities requiring extended focus)	101 vs. 224	15.86	0.67	< 0.001				
Scale 3 (Modern devices and communication skills)	100 vs. 224	22.32	0.98	< 0.001				
Scale 4 (Administrative tasks)	101 vs. 224	17.92	0.73	< 0.001				
Scale 5 (Social activities)	98 vs. 222	19.86	0.67	< 0.001				
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	88 vs. 204	20.12	0.54	< 0.001				
Single-item scale 2 (Difficulty managing own medication)	101 vs. 222	23.33	0.77	< 0.001				
IADL sum score	102 vs. 224	25.12	1.17	< 0.001				
ΔBI ≥ − 10 points vs. < − 10 points								
Scale 1 (Domestic activities)					28 vs. 205	− 18.59	− 0.82	< 0.001
Scale 2 (Activities requiring extended focus)					29 vs. 206	− 13.07	− 0.71	0.04
Scale 3 (Modern devices and communication skills)					28 vs. 206	− 9.60	− 0.53	-
Scale 4 (Administrative tasks)					26 vs. 206	− 13.09	− 0.63	-
Scale 5 (Social activities)					26 vs. 203	− 13.87	− 0.54	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					21 vs. 183	− 24.98	− 0.73	0.01
Single-item scale 2 (Difficulty managing own medication)					29 vs. 204	− 19.86	− 0.75	< 0.05
IADL sum score					29 vs. 206	− 15.87	− 0.93	< 0.01
ΔBI ≥ + 10 points vs. < + 10 points								
Scale 1 (Domestic activities)					9 vs. 244	− 12.16	− 0.52	-
Scale 2 (Activities requiring extended focus)					9 vs. 226	6.34	0.34	-
Scale 3 (Modern devices and communication skills)					9 vs. 225	1.26	0.07	-
Scale 4 (Administrative tasks)					9 vs. 223	0.51	0.02	-
Scale 5 (Social activities)					9 vs. 220	4.89	0.19	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					8 vs. 198	4.76	0.14	-
Single-item scale 2 (Difficulty managing own medication)					8 vs. 225	21.43	0.79	-
IADL sum score					9 vs. 235	− 1.51	− 0.08	-
Neurocognitive status (NCS)								
Not neurocognitively impaired vs. neurocognitively impaired								
Scale 1 (Domestic activities)	202 vs. 122	− 15.66	− 0.51	< 0.001				
Scale 2 (Activities requiring extended focus)	201 vs. 124	− 17.60	− 0.75	< 0.001				
Scale 3 (Modern devices and communication skills)	201 vs. 123	− 23.72	− 1.07	< 0.001				
Scale 4 (Administrative tasks)	201 vs. 124	− 22.27	− 0.95	< 0.001				
Scale 5 (Social activities)	199 vs. 121	− 17.69	− 0.60	< 0.001				
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	189 vs. 103	− 16.24	− 0.43	< 0.001				
Single-item scale 2 (Difficulty managing own medication)	202 vs. 121	− 20.19	− 0.66	< 0.001				
IADL sum score	202 vs. 124	− 18.32	− 0.81	< 0.001				
ΔNCS ↓ vs. ↑/=								
Scale 1 (Domestic activities)					22 vs. 211	− 27.21	− 1.18	< 0.001
Scale 2 (Activities requiring extended focus)					23 vs. 212	− 17.73	− 0.92	< 0.001
Scale 3 (Modern devices and communication skills)					23 vs. 211	− 24.30	− 1.36	< 0.001
Scale 4 (Administrative tasks)					21 vs. 211	− 23.90	− 1.13	< 0.001
Scale 5 (Social activities)					22 vs. 207	− 15.66	− 0.60	0.02
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					18 vs. 186	− 18.82	− 0.53	0.02
Single-item scale 2 (Difficulty managing own medication)					22 vs. 211	− 19.60	− 0.72	0.04
IADL sum score					23 vs. 212	− 23.09	− 1.31	< 0.001
ΔNCS ↑ vs. ↓/=								
Scale 1 (Domestic activities)					18 vs. 215	1.15	0.05	-
Scale 2 (Activities requiring extended focus)					18 vs. 217	3.11	0.16	-
Scale 3 (Modern devices and communication skills)					17 vs. 217	1.68	0.09	-
Scale 4 (Administrative tasks)					17 vs. 215	− 0.51	− 0.02	-
Scale 5 (Social activities)					18 vs. 211	1.02	0.04	-
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					18 vs. 186	12.72	0.36	-
Single-item scale 2 (Difficulty managing own medication)					18 vs. 215	3.55	0.13	-
IADL sum score					18 vs. 217	2.08	0.11	-
Subjective neurocognitive complaints (MOS CFS-R)								
MOS CFS-R ≤ 30 vs. MOS CFS-R > 30								
Scale 1 (Domestic activities)	170 vs. 141	22.57	0.78	< 0.001				
Scale 2 (Activities requiring extended focus)	171 vs. 141	28.06	1.41	< 0.001				
Scale 3 (Modern devices and communication skills)	170 vs. 141	25.29	1.20	< 0.001				
Scale 4 (Administrative tasks)	171 vs. 141	26.96	1.26	< 0.001				
Scale 5 (Social activities)	168 vs. 138	30.74	1.16	< 0.001				
Single-item scale 1 (Difficulty doing your job (paid or voluntary)	149 vs. 134	28.92	0.81	< 0.001				
Single-item scale 2 (Difficulty managing own medication)	168 vs. 141	22.35	0.75	< 0.001				
IADL sum score	171 vs. 141	25.81	1.30	< 0.001				
ΔMOS CFS-R ≥ − 6 points vs. < − 6 points								
Scale 1 (Domestic activities)					25 vs. 199	27.34	1.22	< 0.001
Scale 2 (Activities requiring extended focus)					26 vs. 199	23.01	1.24	< 0.001
Scale 3 (Modern devices and communication skills)					25 vs. 199	20.85	1.15	< 0.001
Scale 4 (Administrative tasks)					24 vs. 199	30.40	1.49	< 0.001
Scale 5 (Social activities)					25 vs. 196	26.54	1.08	< 0.001
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					20 vs. 177	40.73	1.23	< 0.001
Single-item scale 2 (Difficulty managing own medication)					25 vs. 198	26.86	1.00	< 0.001
IADL sum score					26 vs. 199	25.96	1.55	< 0.001
ΔMOS CFS-R ≥ + 6 points vs. < + 6 points								
Scale 1 (Domestic activities)					23 vs. 200	− 21.60	− 0.93	< 0.001

(*continued on next page*)

Table 3 (*continued*)

	Baseline				*Over time*			
Sociodemographic	*N*	*MD*	*d*	*p*	*N*	*MD*	*d*	*p*
Scale 2 (Activities requiring extended focus)					23 vs. 201	− 17.96	− 0.93	< 0.001
Scale 3 (Modern devices and communication skills)					23 vs. 200	− 13.72	− 0.73	< 0.01
Scale 4 (Administrative tasks)					23 vs. 199	− 20.47	− 0.95	< 0.001
Scale 5 (Social activities)					22 vs. 198	− 16.78	− 0.65	< 0.01
Single-item scale 1 (Difficulty doing your job (paid or voluntary)					19 vs. 177	− 3.71	− 0.11	-
Single-item scale 2 (Difficulty managing own medication)					23 vs. 199	− 14.72	− 0.53	0.03
IADL sum score					23 vs. 201	− 18.61	− 1.04	< 0.001

Known-groups analyses was examined using Mann Whitney U tests to determine if differences between groups were significant. Only levels of significance, i.e., $p < 0.05$, are presented. Abbr. N = number, M = Mean, d = Cohen d' (effect size), p = P value (level of significance), yo = years old, LGG = low grade glioma, HGG = high grade glioma, BM = brain metastases, Δ = delta (i.e. difference between groups in change scores between baseline and the 3-month follow-up assessment), KPS = Karnofsky Performance Status, BI = Barthel Index, NCI = neurocognitively impaired, NCS = neurocognitive status, ↓ = decrease in neurocognitive status, ↑/= = increase or stable in neurocognitive status, ↑ = increase in neurocognitive status, ↓/= = decrease or stable in neurocognitive status, MOS CFS-R = MOS Cognitive Functioning Scale-Revised.

Table 4
Test-retest reliability, based on the baseline (T0) and 2-week follow-up assessment (T1), for all EORTC IADL-BN32 scales and the IADL sum score.

	N	T0 Mean (SD)	T1 Mean (SD)	Mean diff (SD)	ICC (CI-95 %)	SEM (%)	SDC (%)
Scale 1: Domestic activities	275	26.4 (29.6)	26.2 (31.3)	0.17 (21.0)	0.76 (0.71–0.81)	1.26 (1 %)	3.49 (3 %)
Scale 2: Activities requiring extended focus	275	24.0 (23.9)	22.7 (23.7)	1.31 (16.6)	0.76 (0.70–0.80)	1.00 (1 %)	2.77 (3 %)
Scale 3: Modern devices and communication skills	275	18.0 (23.5)	18.2 (25.7)	− 0.20 (17.6)	0.74 (0.69–0.79)	1.06 (1 %)	2.94 (3 %)
Scale 4: Administrative tasks	275	16.6 (23.5)	16.4 (25.2)	0.20 (17.5)	0.74 (0.69–0.79)	1.05 (1 %)	2.91 (3 %)
Scale 5: Social activities	268	23.7 (29.5)	24.0 (29.5)	− 0.31 (22.2)	0.72 (0.66–0.77)	1.35 (1 %)	3.74 (4 %)
Single-item scale 1: difficulty doing your job (paid or voluntary)	241	28.9 (36.5)	31.8 (38.6)	− 2.90 (29.0)	0.70 (0.63–0.76)	1.87 (2 %)	5.18 (5 %)
Single-item scale 2: difficulty managing own medication	275	16.2 (29.4)	13.3 (25.3)	2.91 (25.3)	0.57 (0.49–0.65)	1.53 (2 %)	4.24 (4 %)
IADL sum score	276	23.2 (22.7)	22.8 (24.0)	0.37 (14.7)	0.80 (0.76–0.84)	0.88 (1 %)	2.44 (2 %)

Abbr. N = number, SD = standard deviation, Mean diff = mean difference, ICC = intra-class correlation coefficients, SEM = standard error of measurement, SDC = smallest detectable change for individual subjects.

that these differences are caused by differences in sample sizes and patient characteristics between geographic regions. Patients from the Northern European region were overrepresented in the sample, and there were significant differences in baseline sociodemographic and clinical characteristics between the regions. With regards to the construct validity of the EORTC IADL-BN32, the questionnaire was determined to be good at discriminating between patient groups based on variables related to functioning in daily life (performance status and BADL) and neurocognitive functioning (assessed by a HCP (neurocognitive status) and by the patients themselves (neurocognitive complaints)). We considered these functional variables the most important indicators of support for construct validity as they are considered to reflect the aspects of the underlying construct of IADL. Other patient characteristics, such as age and gender, and tumour-characteristics were less able to discriminate between patient's IADL performance. This discriminatory ability of the subgroups was mainly detectable for most variables at the baseline measurement and for patients showing deterioration over time, while it was less obvious for improvements over time. Perhaps the reason that the EORTC IADL-BN32 is less accurate if a patients is regarded as improved rather than deteriorated, is that these two states might be related to two different underlying factors. In general, decline in (neurocognitive) functioning is most likely more observable to the healthcare professional (HCP) and a patient or proxy than improvement in functioning. Since neurocognitive functioning was assessed subjectively in most cases, as is the assessment of IADL with the EORTC IADL-BN32, changes in neurocognitive status and IADL are expected to be (strongly) correlated. Also, the evaluation of the HCP with respect to the performance and neurocognitive status of the patient, which is determined in a short time period during a consultation, might not be congruent with the experience of the patient in their day-to-day life. This is supported by the fact that we do find significant effects when looking at improvements in both IADL and subjective cognitive complaints, with this latter variable also reflecting the perspective of the patient. We detected a similar pattern in the responsiveness of the EORTC IADL-BN32. The questionnaire showed reasonable responsiveness when it came to deterioration on other functional measures, and less responsiveness when it came to improvements on other functional measures, except, again, for subjective neurocognitive complaints. However, results of the responsiveness analyses were based on subgroups with small sample sizes and therefore should be confirmed in a study with larger sample sizes for the subgroups. Lastly, the test-retest reliability was good, with the exception of Single-item scale 2 (*Difficulty managing own medication*). This indicates that patients seem to experience rapid changes in their ability to manage their medication, or find it difficult to report on this ability, and is therefore a less consistent outcome.

4.1. Limitations and future studies

Although we aimed to recruit a similar number of patients in the different geographical regions, recruitment was severely impacted by the COVID-19 pandemic, which limited the number of centres that could recruit patients, as well as access to patients in centres open for recruitment. Nevertheless, the distribution of tumour types and neurocognitive status of the population was well-balanced at baseline, and the compliance rate at the different assessment times was high.

The distribution of the tumour types was based on the WHO 2016 classification, as this was the gold standard at the time this study was designed and conducted. Currently, the WHO 2021 classification is available. According to the new classification 14/220 glioma patients would be reclassified as the IDH-mutant anaplastic astrocytoma and anaplastic oligodendroglioma patients would now be classified, as a low-grade glioma instead of a high-grade/fast growing tumour. This could have influenced the results. In addition, it is a major limitation that the neurocognitive status of the patients was not assessed with an objective test battery in most cases (80 %). Based on the results of phase 3 of the development of the IADL questionnaire, performing a neuropsychological assessment for all patients included in this study was not deemed feasible in this phase 4 study as this was considered too time consuming for the participating centres and too burdensome for the

Table 5
Responsiveness of the EORTC IADL-BN32 questionnaire. The sensitivity to changes in IADL outcomes between baseline (T0) and the 3-month follow-up (T2) is presented for different subgroups of patients.

		N	T0 Mean (SD)	T2 Mean (SD)	Mean diff. (SD)	p	SMD	SRM
Scale 1: Domestic activities	All patients	236	25.2 (28.8)	29.7 (31.4)	− 4.48 (24.6)	-	− 0.16	− 0.18
	KPS deteriorated (≥ 10 points)	46	27.0 (31.6)	44.9 (37.1)	− 17.90 (34.0)	< 0.01	− 0.57	− 0.53
	KPS improved (≥ 10 points)	38	34.4 (31.6)	38.4 (35.2)	− 4.02 (23.2)	-	− 0.13	− 0.17
	BI deteriorated (≥ 10 points)	28	48.8 (35.1)	69.0 (35.8)	− 20.16 (29.9)	< 0.01	− 0.57	− 0.67
	BI improved (≥ 10 points)	9	53.5 (30.4)	69.0 (33.1)	− 15.49 (38.4)	-	− 0.51	− 0.40
	Neurocognitive status deteriorated	22	33.9 (32.0)	62.8 (32.4)	− 28.88 (36.9)	< 0.01	− 0.90	− 0.78
	Neurocognitive status improved	18	35.1 (29.9)	40.4 (30.3)	−5.30 (24.4)	-	− 0.18	− 0.22
	Subj. neurocogn. complaints increased (≥ 6 points)	25	33.9 (32.1)	61.9 (35.9)	− 28.04 (38.8)	< 0.01	− 0.87	− 0.72
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	48.7 (34.7)	33.2 (31.3)	15.59 (24.7)	< 0.01	0.45	0.63
Scale 2: Activities requiring extended focus	All patients	238	24.4 (24.0)	25.6 (24.5)	− 1.22 (20.2)	-	− 0.05	− 0.06
	KPS deteriorated (≥ 10 points)	46	21.4 (20.8)	32.52 (29.7)	− 11.09 (27.3)	0.02	− 0.53	− 0.41
	KPS improved (≥ 10 points)	38	32.5 (28.0)	29.9 (24.4)	2.52 (23.4)	-	0.09	0.11
	BI deteriorated (≥ 10 points)	29	35.0 (25.0)	47.0 (31.3)	− 12.01 (29.4)	-	− 0.48	− 0.41
	BI improved (≥ 10 points)	9	43.5 (37.7)	38.0 (32.1)	5.56 (38.2)	-	− 0.15	− 0.15
	Neurocognitive status deteriorated	23	33.7 (29.9)	50.7 (30.1)	−17.05 (34.6)	0.02	− 0.57	− 0.49
	Neurocognitive status improved	18	36.6 (26.0)	40.5 (29.0)	− 3.94 (21.1)	-	− 0.15	− 0.19
	Subj. neurocogn. complaints increased (≥ 6 points)	26	30.3 (26.4)	51.7 (29.7)	− 21.38 (34.0)	< 0.01	− 0.81	− 0.63
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	43.5 (30.9)	28.4 (22.2)	15.04 (21.6)	< 0.01	0.49	0.70
Scale 3: Modern devices and communication skills	All patients	235	18.0 (23.5)	18.2 (25.7)	− 0.20 (17.6)	-	− 0.05	− 0.01
	KPS deteriorated (≥ 10 points)	46	16.5 (21.7)	25.4 (32.3)	− 8.91 (29.6)	-	− 0.41	− 0.30
	KPS improved (≥ 10 points)	38	25.2 (26.5)	20.7 (25.9)	4.52 (17.2)	-	0.17	0.26
	BI deteriorated (≥ 10 points)	28	26.8 (24.7)	35.9 (36.1)	− 9.11 (33.7)	-	− 0.37	− 0.27
	BI improved (≥ 10 points)	9	28.9 (34.2)	28.3 (34.2)	0.56 (31.8)	-	0.02	0.02
	Neurocognitive status deteriorated	23	24.9 (23.6)	48.1 (35.9)	− 23.19 (34.5)	< 0.01	− 0.98	− 0.67
	Neurocognitive status improved	17	33.7 (28.8)	36.6 (27.6)	− 2.84 (18.4)	-	− 0.10	− 0.15
	Subj. neurocogn. complaints increased (≥ 6 points)	25	22.7 (25.4)	42.2 (36.4)	− 19.53 (33.9)	< 0.01	− 0.77	− 0.58
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	27.2 (30.1)	16.0 (25.0)	11.23 (20.0)	0.01	0.37	0.56
Scale 4: Administrative tasks	All patients	235	16.6 (23.4)	16.0 (24.4)	0.64 (22.5)	-	0.03	0.03
	KPS deteriorated (≥ 10 points)	45	16.3 (25.0)	25.6 (32.6)	− 9.26 (32.7)	-	− 0.37	− 0.28
	KPS improved (≥ 10 points)	38	23.7 (30.4)	17.5 (29.1)	6.14 (25.0)	0.03	0.20	0.25
	BI deteriorated (≥ 10 points)	26	21.4 (29.6)	31.6 (35.8)	− 10.26 (34.3)	-	− 0.35	− 0.30
	BI improved (≥ 10 points)	9	35.8 (37.2)	34.0 (42.0)	1.85 (42.1)	-	0.05	0.04
	Neurocognitive status deteriorated	21	24.9 (25.8)	45.8 (35.6)	− 20.90 (37.7)	0.02	− 0.81	− 0.55
	Neurocognitive status improved	17	36.6 (31.9)	35.3 (27.8)	1.31 (20.4)	-	0.04	0.06
	Subj. neurocogn. complaints increased (≥ 6 points)	24	20.8 (27.7)	47.0 (35.2)	− 26.16 (36.9)	< 0.01	− 0.94	− 0.71
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	34.3 (32.5)	15.0 (20.0)	19.32 (24.4)	< 0.01	0.59	0.79
Scale 5: Social activities	All patients	232	24.3 (29.9)	27.3 (31.0)	− 3.02 (27.14)	-	− 0.10	− 0.11
	KPS deteriorated (≥ 10 points)	43	20.0 (26.0)	35.9 (33.9)	− 15.89 (34.1)	< 0.01	− 0.61	− 0.47
	KPS improved (≥ 10 points)	38	37.4 (39.4)	33.2 (34.6)	4.24 (26.8)	-	0.11	0.16
	BI deteriorated (≥ 10 points)	26	32.9 (29.7)	47.4 (36.2)	− 14.53 (31.4)	0.04	− 0.49	− 0.46
	BI improved (≥ 10 points)	9	43.2 (44.9)	40.7 (36.0)	2.47 (51.2)	-	0.06	0.05

(*continued on next page*)

Table 5 (*continued*)

		N	T0 Mean (SD)	T2 Mean (SD)	Mean diff. (SD)	p	SMD	SRM
	Neurocognitive status deteriorated	22	34.8 (37.0)	51.8 (34.9)	− 16.92 (41.2)	-	− 0.46	− 0.41
	Neurocognitive status improved	18	45.7 (38.3)	49.4 (34.1)	− 3.70 (32.3)	-	− 0.10	− 0.11
	Subj. neurocogn. complaints increased (≥ 6 points)	25	29.3 (34.2)	55.3 (33.7)	− 26.00 (41.6)	< 0.01	− 0.76	− 0.63
	Subj. neurocogn. complaints decreased (≥ 6 points)	22	43.4 (36.6)	30.8 (27.2)	12.63 (28.7)	-	0.35	0.44
Single-item scale 1: difficulty doing your job (paid or voluntary)	All patients	206	28.6 (36.4)	33.7 (38.5)	− 5.02 (35.5)	-	− 0.14	− 0.14
	KPS deteriorated (≥ 10 points)	37	25.2 (34.6)	42.3 (36.6)	− 17.12 (37.4)	0.02	− 0.49	− 0.46
	KPS improved (≥ 10 points)	35	39.0 (44.6)	41.9 (43.8)	− 2.86 (41.5)	-	− 0.06	− 0.07
	BI deteriorated (≥ 10 points)	21	49.2 (44.2)	76.2 (36.7)	− 26.99 (53.4)	0.03	− 0.61	− 0.51
	BI improved (≥ 10 points)	8	54.2 (50.2)	54.2 (35.4)	0.00 (39.8)	-	0.00	0.00
	Neurocognitive status deteriorated	18	37.0 (36.0)	59.3 (34.0)	− 22.22 (39.6)	0.03	− 0.62	− 0.56
	Neurocognitive status improved	18	38.9 (34.8)	55.6 (37.9)	− 16.67 (34.8)	< 0.05	− 0.48	− 0.48
	Subj. neurocogn. complaints increased (≥ 6 points)	20	38.3 (40.9)	80.0 (33.2)	− 41.67 (48.2)	< 0.01	− 1.02	− 0.86
	Subj. neurocogn. complaints decreased (≥ 6 points)	19	47.4 (47.6)	49.1 (42.1)	− 1.75 (32.3)	-	− 0.04	− 0.05
Single-item scale 2: difficulty managing own medication	All patients	224	11.8 (24.0)	15.5 (29.3)	− 3.72 (22.6)	-	− 0.02	− 0.17
	KPS deteriorated (≥ 10 points)	45	15.6 (27.2)	31.1 (37.9)	− 15.56 (37.3)	0.01	− 0.57	− 0.42
	KPS improved (≥ 10 points)	38	16.7 (30.8)	11.4 (24.8)	5.26 (27.4)	-	0.17	0.19
	BI deteriorated (≥ 10 points)	29	24.1 (29.4)	41.4 (43.3)	− 17.24 (45.1)	-	− 0.59	− 0.38
	BI improved (≥ 10 points)	8	45.8 (46.9)	25.0 (34.5)	20.83 (50.2)	-	0.44	0.41
	Neurocognitive status deteriorated	22	25.8 (34.0)	43.9 (42.9)	−18.18 (52.2)	-	− 0.53	− 0.35
	Neurocognitive status improved	18	25.9 (37.1)	29.6 (32.1)	− 3.70 (30.0)	-	− 0.10	− 0.12
	Subj. neurocogn. complaints increased (≥ 6 points)	25	25.3 (33.7)	49.3 (45.3)	− 24.00 (50.5)	0.03	− 0.71	− 0.48
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	36.2 (38.8)	23.2 (35.4)	13.04 (35.9)	-	0.34	0.36
IADL sum score	All patients	226	21.4 (22.7)	24.5 (24.3)	− 3.13 (16.8)	-	− 0.11	− 0.19
	KPS deteriorated (≥ 10 points)	46	22.0 (22.2)	36.0 (30.4)	− 14.01 (27.7)	< 0.01	− 0.63	− 0.51
	KPS improved (≥ 10 points)	38	31.3 (25.0)	30.3 (24.5)	1.05 (18.5)	-	0.04	0.06
	BI deteriorated (≥ 10 points)	29	37.3 (22.3)	53.1 (29.3)	− 15.79 (26.5)	< 0.01	− 0.71	− 0.60
	BI improved (≥ 10 points)	9	44.0 (30.5)	47.3 (28.4)	− 3.32 (36.4)	-	− 0.11	− 0.09
	Neurocognitive status deteriorated	23	31.7 (26.1)	54.9 (28.8)	− 23.21 (33.1)	< 0.01	− 0.89	− 0.70
	Neurocognitive status improved	18	36.9 (27.4)	41.2 (28.8)	− 4.30 (19.8)	-	− 0.16	− 0.22
	Subj. neurocogn. complaints increased (≥ 6 points)	26	30.2 (24.9)	55.2 (28.2)	− 25.02 (31.4)	< 0.001	− 1.00	− 0.80
	Subj. neurocogn. complaints decreased (≥ 6 points)	23	42.0 (26.5)	27.4 (21.7)	14.59 (17.6)	< 0.01	0.55	0.83

Abbr. N = number, SD = standard deviation, p = P value (level of statistical significance), SMD = standardized mean difference, SRM = standardized response mean, KPS = Karnofsky Performance Status, BI = Barthel Index, Subj. neurocogn. = subjective neurocognitive.

patients (particularly those with poor health). We therefore choose other (suboptimal) sources of information on the patient's neurocognitive status. In clinical practice, patients are often assessed using clinical outcomes determined by the treating physician (i.e. KPS and BI). However, future studies could validate these findings by reproducing the results by including a neuropsychological assessment for all patients to determine neurocognitive status more objectively.

Although the overall study population is representative of the brain tumour population as a whole based on tumour type and neurocognitive status, the differences in sample sizes between regions presumably did affect the initial results related to cross-cultural validity. Particularly patients from Northern Europe showed some item response biases, in comparison to patients from the Southern European and Non-European regions. However, this analysis may have been hampered as a number of countries were merged together into smaller categories for this analysis. The results should therefore be interpreted with caution, as the questionnaire may not be sufficiently valid in all countries, separate cross-cultural validation of the EORTC IADL-BN32 for their patient population is therefore encouraged, and further research is warranted before implementation in clinical studies or clinical practice. Furthermore, larger longitudinal studies are needed to confirm the results based on the smaller sample sizes analysed over time between the baseline and 3-

month FU assessments. Future studies should determine thresholds to detect clinically meaningful change over time for the EORTC IADL-BN32, which is considered an important aspect when interpreting patient-reported outcomes by the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI) Consortium [32].

Lastly, the 'not applicable' (NA) option was not included in the phase IV version of the EORTC IADL-BN32. The NA option was included in phase III of the project [14], but resulted in difficulties with the statistical analyses. We therefore recommend that patients are encouraged to provide an indication on their level of performance on the different IADLs, even if they did not perform a specific activity. Respondents are encouraged to fill in no difficulties (i.e., 'not at all') if they estimated that they could perform the activity without issues or 'very much' if they would not be able to perform that IADL at all, e.g., due to poor health.

4.2. Clinical implications

Measuring IADL can give insight in an important aspect of patients functioning, i.e., the level of independence in daily life. In addition, the questionnaire may be used to obtain information on the daily life implications of neurocognitive status of a patient, as patients with neurocognitive deficits or complaints in general report more problems with IADL. The information obtained with the EORTC IADL-BN32 can be used in clinical practice to monitor patients' functioning over time, and implement supportive treatments (e.g., rehabilitation) if needed. Furthermore, the EORTC IADL-BN32 questionnaire would be an important endpoint in clinical trials to gain insight in the impact of treatments on the patients' level of functioning.

For now, we suggest adopting the patient-version of the EORTC IADL-BN32 if the aim is to assess IADL in clinical studies or practice. The proxy-version of the EORTC IADL-BN32 or the EORTC QLQ-C30 and QLQ-BN20 questionnaires could be used in conjunction with the EORTC IADL-BN32, depending on the research or clinical question. However, as we expect neurocognitive impairments could play a role in the accuracy of the self-reported IADL assessment by patients, it is recommended to additionally assess the proxy-version of the EORTC IADL-BN32 as it could provide additional information on the patients' functioning in daily life and/or include a neuropsychological test battery, particularly in case the patient has neurocognitive deficits. This information would give insight into if and when, the proxy-version might be valuable in addition to, or even instead of the patient-version of the IADL-BN32. As reported in one of our previous studies [33], proxies did report more problems on the EORTC IADL-BN32 questionnaire in general, but this effect was more apparent in dyads with a neurocognitively impaired patient. Currently, we are analysing the proxy data collected during this international validation study to get more insight in this matter.

4.3. Conclusion

In summary, this international phase IV validation has shown that the EORTC IADL-BN32 questionnaire has adequate to excellent psychometric properties and can be used in clinical trials and practice for patients with brain tumours. Several aspects, such as the cross-cultural validity and responsiveness (including establishing clinically important changes), are not yet optimal and should be further explored.

CRediT authorship contribution statement

Duska Petranovic: Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Maja Ploh:** Writing – review & editing, Validation, Investigation, Data curation. **Waleed Alrjoob:** Writing – review & editing, Methodology, Investigation, Data curation. **Andrea Pace:** Writing – review & editing, Validation, Investigation, Data curation. **Joana Silva:** Writing – review & editing, Validation, Investigation. **Marianne Hjermstad:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Andreia Capela:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Nadja Talhi:** Writing – review & editing, Investigation, Conceptualization. **Jaap Reijneveld:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Josef Pichler:** Writing – review & editing, Validation, Methodology, Data curation. **Sietske Sikkes:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Tadeja Urbanic Purkart:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Clemens Seidel:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Quirien Oort:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Melanie Hand:** Writing – review & editing, Investigation, Data curation. **Teresa Young:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Christine Brannan:** Writing – review & editing, Validation, Investigation, Data curation. **Isolde Höllmüller:** Writing – review & editing, Validation, Investigation, Data curation. **Johan Koekkoek:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Lorna Brown:** Writing – review & editing, Investigation, Data curation. **Florien Boele:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Tara Chalk:** Writing – review & editing, Data curation. **Bernard Uitdehaag:** Writing – review & editing, Project administration, Funding acquisition. **Andrea Talacchi:** Writing – review & editing, Validation, Investigation, Data curation, Conceptualization. **Martin Taphoorn:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Annapina Mazzotta:** Writing – review & editing, Investigation, Data curation. **Martin Klein:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Neil Aaronson:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Yasuji Miyakita:** Writing – review & editing, Validation, Investigation, Data curation. **Omar Shamieh:** Writing – review & editing, Methodology, Data curation. **Linda Dirven:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yoshitaka Narita:** Writing – review & editing, Validation, Investigation, Data curation. **Hitomi Sato:** Writing – review & editing, Validation, Investigation, Data curation.

Ethics approval

Ethical and research governance approvals were obtained at each participating centre in accordance with local requirements. Ethical approval: Dutch Medical Ethical committee (reference no.: 2013.289); London-Bloomsbury Research Ethics Committee REC (reference no.: 14/LO/0452); South East Scotland Research Ethics Committee, Edinburgh IRAS (Project ID 148706); Italian Comitato Etico per la Sperimentazione clinica delle Province di Verona e Rovigo (CESC) (Prog. 758CECS); Ethics Committee of the Medical University of Innsbruck (EC no. 1069/2017); National Cancer Center Institutional Review Board (registration no.: 2017-166). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee mentioned above and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Funding

This study was funded by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group, with Grant no. 005_2019. The EORTC QLG business model involves license fees for commercial use of their instruments. Academic use of EORTC instruments is free of charge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Study data can be requested via the Data Sharing Policy of the European Organisation for Research and Treatment of Cancer (EORTC): https://www.eortc.org/data-sharing. Moreover, the EORTC IADL-BN32 questionnaire can be requested through the website of the EORTC Quality of Life Group: https://qol.eortc.org/questionnaires/.

Acknowledgment

The authors want to sincerely thank all patients, their proxies and health care professionals for their willingness to participate and valuable input in the development of EORTC IADL-BN32 questionnaire. Furthermore, we would like to thank all the health care professionals from each centre for their help with the recruitment of the patients and proxies, and thank the following for their help with patient and proxy data collection in particular.

Informed consent

All participants provided written informed consent before participation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejca.2024.114335.

References

[1] Buckner JC, Brown PD, O'Neill BP, Meyer FB, Wetmore CJ, Uhm JH. Central nervous system tumors. Mayo Clin Proc 2007;82(10):1271–86.
[2] American Association of Neurological Surgons. Brain tumors. ⟨https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors⟩.
[3] Porter KR, McCarthy BJ, Berbaum ML, Davis FG. Conditional survival of all primary brain tumor patients by age, behavior, and histology. Neuroepidemiology 2011;36(4):230–9. https://doi.org/10.1159/000327752. Epub 2011 Jun 16.
[4] Davies E, Clarke C, Hopkins A. Malignant cerebral glioma–I: survival, disability, and morbidity after radiotherapy. BMJ 1996;313(7071):1507–12.
[5] Osoba D, Brada M, Prados MD, Yung WK. Effect of disease burden on health-related quality of life in patients with malignant gliomas. Neuro Oncol 2000;2(4):221–8.
[6] Taphoorn MJ, Schiphorst AK, Snoek FJ, et al. Cognitive functions and quality of life in patients with low-grade gliomas: the impact of radiotherapy. Ann Neurol 1994;36(1):48–54. https://doi.org/10.1002/ana.410360111.
[7] Tucha O, Smely C, Preier M, Lange KW. Cognitive deficits before treatment among patients with brain tumors. Neurosurgery 2000;47(2):324–33. https://doi.org/10.1097/00006123-200008000-00011 [discussion 33-4].
[8] Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement 2015;11(10):1231–40. https://doi.org/10.1016/j.jalz.2014.10.006. Epub 5 Jan 15.
[9] Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement 2012;8(6):536–43. https://doi.org/10.1016/j.jalz.2011.08.006.
[10] Sikkes SA, Knol DL, Pijnenburg YA, de Lange-de Klerk ES, Uitdehaag BM, Scheltens P. Validation of the Amsterdam IADL Questionnaire(c), a new tool to measure instrumental activities of daily living in dementia. Neuroepidemiology 2013;41(1):35–41. https://doi.org/10.1159/000346277. Epub 2013 May 25.
[11] Sikkes SA, Pijnenburg YA, Knol DL, de Lange-de Klerk ES, Scheltens P, Uitdehaag BM. Assessment of instrumental activities of daily living in dementia: diagnostic value of the Amsterdam Instrumental Activities of Daily Living Questionnaire. J Geriatr Psychiatry Neurol 2013;26(4):244–50. https://doi.org/10.1177/0891988713509139.
[12] Oort Q, Dirven L, Meijer W, et al. Development of a questionnaire measuring instrumental activities of daily living (IADL) in patients with brain tumors: a pilot study. J Neurooncol 2017;132(1):145–53. https://doi.org/10.1007/s11060-016-2352-1. Epub 2017 Feb 1.
[13] EORTC module guidelines fifth edition. EORTC QOL Group; 2021.
[14] Oort Q, Dirven L, Sikkes SAM, et al. Development of an EORTC questionnaire measuring instrumental activities of daily living (IADL) in patients with brain tumours: phase I–III. Qual Life Res 2021;30(5):1491–502.
[15] EORTC QLQ-C30 scoring manual. EORTC QOL Group; 2001.
[16] Hays RD, Sherbourne CD, Mazel R. User's manual for the medical outcomes study (MOS) core measures of health-related quality of life. Santa Monica, CA: RAND Corporation; 1995.
[17] Karnofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of the nitrogen mustards in the palliative treatment of carcinoma: with particular reference to bronchogenic carcinoma. Cancer 1948.
[18] Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J 1965;14:61–5.
[19] Kline P . The handbook of psychological testing/P. Kline; 2000.
[20] Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21(2):230–58.
[21] Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model: A Multidiscip J 1999;6(1):1–55.
[22] Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. John Wiley & Sons; 2013.
[23] Holland W, Thayer DT. Differential item performance and the Mantel-Haenszel procedure. Test Validity 1988:129–45.
[24] Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. J Educ Meas 1990;27:361–70.
[25] Lord FM. Applications of item response theory to practical testing problems. Routledge; 1980.
[26] Tabira T, Hotta M, Murata M, et al. Age-related changes in instrumental and basic activities of daily living impairment in older adults with very mild Alzheimer's disease. Dement Geriatr Cogn Disord Extra 2020;10(1):27–37.
[27] Cohen J . Statistical power analysis for the behavioral sciences; 1988.
[28] Abernethy AP, Shelby-James T, Fazekas BS, Woods D, Currow DC. The Australia-modified Karnofsky Performance Status (AKPS) scale: a revised scale for contemporary palliative care clinical practice [ISRCTN81117481]. BMC Palliat Care 2005;4(1):7.
[29] Hsieh Y-W, Wang C-H, Wu S-C, Chen P-C, Sheu C-F, Hsieh C-L. Establishing the minimal clinically important difference of the Barthel index in stroke patients. Neurorehab Neural Repair 2007;21(3):233–8.
[30] Gehring K, Taphoorn MJB, Sitskoorn MM, Aaronson NK. Predictors of subjective versus objective cognitive functioning in patients with stable grades II and III glioma. Neuro-Oncol Pract 2015;2(1):20–31.
[31] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63.
[32] Pe M, Alanya A, Falk RS, et al. Setting international standards in analyzing patient-reported outcomes and quality of life endpoints in cancer clinical trials-innovative medicines initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. Lancet Oncol 2023;24(6):e270–83.
[33] Oort Q, Dirven L, Sikkes SA, et al. Do neurocognitive impairments explain the differences between brain tumor patients and their proxies when assessing the patient's IADL? Neuro-Oncol Pract 2022;9(4):271–83.