Early measures of cognitive function predict survival in patients with newly diagnosed glioblastoma

Derek R. Johnson, Allison M. Sawyer, Christina A. Meyers, Brian Patrick O’Neill, and Jeffrey S. Wefel

Department of Neurology, Mayo Clinic, Rochester, Minnesota (D.R.J., B.P.O.); Section of Neuropsychology, Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas (A.M.S., C.A.M., J.S.W.)

Cognitive dysfunction is a common manifestation of primary brain tumors. We evaluated the association between early cognitive dysfunction and prognosis in a cohort of patients with newly diagnosed glioblastoma. Ninety-one patients who completed neuropsychological assessment after tumor resection but before further treatment were identified in the MD Anderson Neuropsychology database. The relationship between performance on cognitive testing and survival was evaluated using not only Cox proportional hazards models that included clinical factors such as age and KPS but also the Kaplan–Meier method. Median survival time from surgery was 20.7 months. Rates of impairment on cognitive testing ranged from 7.1% for Similarities, to 60.0% for Hopkins Verbal Learning Test–Revised Total Recall. As continuous variables, the Clinical Trial Battery Composite, Trail Making Test Part B, and Controlled Oral Word Association test were associated with survival. Impairment on the Trail Making Test Part B, Controlled Oral Word Association, Similarities, and Digit Span were associated with mortality. Kaplan–Meier analysis demonstrated the survival impact of these tests on the group as a whole and in select patient subgroups defined by classification by the Radiation Therapy Oncology Group (RTOG) Recursive Partitioning Analysis (RPA). Cognitive impairment as measured by specific neuropsychological tests is independently associated with poor prognosis in patients with newly diagnosed glioblastoma, and this effect remains significant even within patient subgroups defined by RTOG RPA class. Executive function and attention are the cognitive domains most closely associated with prognosis in this analysis.

Keywords: cognition, glioblastoma, prognostic, survival.

Glioblastoma, the most common and aggressive primary brain tumor, has a median survival time of less than 2 years. There is a high degree of interpatient variability in survival, with some patients living only months and others living in excess of 5 years. A variety of patient- and tumor-level characteristics affect the prognosis of patients with glioblastoma. Notable patient-level factors include age, extent of resection, and performance status. Tumor-level predictors of survival include methylation status of O6-methylguanine-DNA methyltransferase (MGMT) and mutation of isocitrate dehydrogenase 1 (IDH1). Of these predictors of prognosis, patient performance status is the most subjective as well as the most multifaceted, reflecting both motor and cognitive dysfunction, either of which can take a number of forms.

The concept of performance status is not unique to neuro-oncology. KPS, the scoring system in most widespread use, was developed for use in patients with non-CNS tumors. The KPS scale classifies patients based on the presence and intensity of symptoms, the ability to work, and the need for assistance with self-care but does not distinguish among different types of symptoms or performance limitations. While the KPS scale has proven useful in stratifying brain tumor patients for clinical trials, it may obscure clinically important differences among patients. For example, a patient with mild cognitive dysfunction may have a different prognosis than a patient with a mild hemiparesis, despite identical KPS scores. This potential issue is magnified by the frequent use of KPS scores within other prognosis...
classification schemas, such as the Radiation Therapy Oncology Group (RTOG) Recursive Partitioning Analysis (RPA) classification. In this analysis, we evaluate the prognostic significance of objective neurocognitive measures obtained soon after diagnosis in patients with glioblastoma.

**Materials and Methods**

**Participants**

Adult patients with lobar glioblastoma who received detailed neuropsychological evaluation after initial tumor resection between 2001 and 2010 were identified in The University of Texas MD Anderson Cancer Center (MDACC) neuropsychology clinical database. The pathological diagnosis of glioblastoma was made by specialist neuropathologists at MDACC. The MDACC Institutional Review Board approved a retrospective analysis. These data were used for use in studies of cognition in patients with cancer.17–21 The CTB Comp was dichotomized using a median split to explore its association with survival in patient groups.

**Data Collection and Coding**

Cognitive testing was conducted by a neuropsychologist or a trained psychometrician supervised by a neuropsychologist. Cognitive test scores were normalized using published normative data8–14 accounting for patient age, gender, and level of education when appropriate, and converted into standardized scores (z-scores; mean = 0, SD = 1). Table 2 lists the cognitive tests routinely performed on brain tumor patients. Standardized assessment with the Hopkins Verbal Learning Test Revised (HVLT-R Delayed Recall [DR]) and HVLT-R Delayed Recognition (Recog) was begun at the midpoint of our study period. Motor test scores (Grip Strength) were normalized after adjusting for age, gender, and handedness using normative data for Caucasian patients with 16–17 years of education.10 Performance on an individual cognitive test was considered impaired if the patient’s z-score was −1.5 or less, in line with standards used in neuropsychological practice. Additionally, 7 of 12 patients who were not able to complete the Trail Making Test B (TMTB) were impaired on the Trail Making Test A (TMTA) during the postsurgical testing session, and these patients were also considered impaired on TMTB for data analysis purposes. The continuous Grip Strength score was the calculated difference between the standardized grip strength scores in the hands contralateral and ipsilateral to the tumor. Grip Strength was considered impaired if the power in the hand contralateral to the tumor was 1.5 or more standard deviations less than the power in the hand ipsilateral to the tumor. As part of the postsurgical cognitive evaluation, patients completed the Functional Assessment of Cancer Therapy–Brain assessment. These data were used to provide information on ability to work, which is necessary for assignment of RTOG RPA class. Patients who agreed with the statement “I am able to work (include work at home)” either “Very much” or “Quite a bit” were considered able to work, while those who replied “Somewhat,” “A little bit,” or “Not at all” were considered unable to work.

All information on patients’ clinical status and treatment came from the medical records, except as noted above. The KPS from the day of cognitive testing was used when available; otherwise, the first postoperative KPS score documented at or after the time of hospital discharge was used for analysis. Information concerning antiepileptic drug and corticosteroid use came from the same clinic visit as the KPS score whenever possible. No distinction was made between stable and tapering doses of corticosteroids. MGMT methylation analysis results were available for only a small subset of patients and thus were not included in analysis. Likewise, many patients were surgically treated at MDACC with postsurgical care elsewhere, so information on salvage treatment at the time of tumor progression is incomplete. When patients were known to have received bevacizumab, either with temozolomide and radiation or at time of progression, this information was noted for inclusion in an exploratory model. Tumor volume was determined for all patients using MRI with MedVision 1.41 software, as previously described.16 Extent of resection was calculated using the preoperative and postoperative tumor volumes.

**Statistical Analysis**

Descriptive statistics for demographic variables were generated with means and SDs or medians with interquartile ranges as appropriate. Pearson χ² tests, t tests, or Wilcoxon tests were used to compare differences in clinical characteristics between selected impaired and unimpaired patient groups.

Survival analysis was performed using Cox proportional hazards models. The initial model included clinical/demographic predictors of outcome but not information on cognitive test results. Subsequent models included the cognitive test of interest, patient age, and KPS given the validated significance of these factors, as well as the interval from resection to evaluation. Cognitive test results were both analyzed as continuous variables and dichotomized by impairment. Additionally, 2 derived composite variables were evaluated. The first was the raw number of impaired tests for an individual patient amongst the tests shown to be significantly associated with survival. The second, referred to as the Clinical Trial Battery Composite (CTB Comp), was the mean of the z-scores for the Controlled Oral Word Association (COWA) test, TMTA, TMTB, HVLT-R Total Recall (TR), HVLT-R DR, and HVLT-R Recog. These tests were chosen in light of their inclusion in a battery that has been recommended for use in studies of cognition in patients with cancer.17–21 The CTB Comp was dichotomized using a median split to explore its association with survival in the Cox models. To control the error rate due to multiple testing, only tests or derived composites associated with survival at the P ≤ .01 level, either as a continuous
predictor or when dichotomized, were considered significant. Cognitive tests significantly associated with survival were further evaluated in exploratory models that also included the volumetric extent of resection, initial treatment (radiation alone vs radiation plus temozolomide, given a change in the standard of care during the course of the study), and confirmed history of treatment with bevacizumab.

For cognitive tests on which impairment was associated with survival, Kaplan–Meier curves were constructed with impairment as the stratification factor. This was done for the patient group as a whole, and within subgroups defined by RTOG RPA class. The log-rank test was used to evaluate statistical significance in Kaplan–Meier testing. All statistical analysis was performed in JMP 9.0 (SAS Institute), and Kaplan–Meier plots were generated in R.22 Two-sided tests with a level of significance of $P \leq .05$ were used, except as noted above.

**Results**

Ninety-one patients with newly diagnosed glioblastoma were evaluated. Patient demographic and baseline characteristics are displayed in Table 1. Overall median survival time was 20.7 months. Median survival was 13.5 months in the subset of patients initially treated with radiation alone and 20.7 months in the subset that received initial treatment with radiation and concurrent temozolomide. At time of analysis, 58 patients were confirmed to have died, and 33 patients were censored. Cognitive test results are displayed in Table 2. Rates of objective impairment ranged from 7.1% for Similarities to 60.0% for HVLT-R TR.

In an initial Cox model that included clinical factors but not cognitive test results, greater patient age ($P < .0001$) was associated with greater risk of death, while the KPS ($P = .9088$) and the interval from resection to evaluation ($P = .3794$) were not. Additionally, there was no relationship between survival and tumor laterality ($P = .3502$) or antiepileptic drug ($P = .7758$) or steroid ($P = .1021$) use, and these factors were not included in any further models. Neither antiepileptic drug nor steroid use was associated with performance on any cognitive test in this cohort (data not shown, $P > .05$ for all tests). Table 3 displays the association between cognitive test results and survival. COWA and CTB Comp were significant predictors of survival as continuous variables, while COWA, TMTB, Digit Span, and Similarities were predictors of survival at the $P < .01$ level when dichotomized by impairment. In the subgroup of patients who received the HVLT-R DR and HVLT-R Recog tests, each test approached but did not meet the $P < .01$ significance threshold. Similarly, dichotomizing the CTB Comp using a median split approached but did not meet the $P < .01$ significance threshold. Sensitivity analyses looking at derived scores for TMT $^{23}$ and Digit Span were also evaluated. There was no association between survival and the continuous or dichotomized TMT ratio score, TMT difference score, Digit Span–Longest Digits Forward, Digit Span–Longest Digits Backward, or Digit Span–Difference Between Longest Digits Forward and Longest Digits Backward (data not shown). The effect of age remained significant in all models, while neither KPS nor time from resection to evaluation was significant in any model. Hazard ratios

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>$N = 91$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>Mean (SD) 53.9 (12.0)</td>
</tr>
<tr>
<td>Sex, % (n)</td>
<td>Male 60 (55)</td>
</tr>
<tr>
<td>Race, % (n)</td>
<td>White 84 (76)</td>
</tr>
<tr>
<td></td>
<td>Hispanic 11 (10)</td>
</tr>
<tr>
<td></td>
<td>Black 4 (4)</td>
</tr>
<tr>
<td></td>
<td>Asian 1 (1)</td>
</tr>
<tr>
<td>Years of education</td>
<td>Mean (SD) 15.2 (2.7)</td>
</tr>
<tr>
<td>Tumor laterality, % (n)</td>
<td>Left 54 (49)</td>
</tr>
<tr>
<td></td>
<td>Right 43 (39)</td>
</tr>
<tr>
<td></td>
<td>Bilateral or multifocal 3 (3)</td>
</tr>
<tr>
<td>Tumor lobe, % (n)</td>
<td>Frontal 41 (37)</td>
</tr>
<tr>
<td></td>
<td>Temporal 44 (40)</td>
</tr>
<tr>
<td></td>
<td>Parietal 13 (12)</td>
</tr>
<tr>
<td></td>
<td>Insula 2 (2)</td>
</tr>
<tr>
<td>Extent of resection</td>
<td>Mean % (SD) 94.1 (12.7)</td>
</tr>
<tr>
<td>Days from resection to testing</td>
<td>Mean (SD) 21.4 (12.7)</td>
</tr>
<tr>
<td>Postoperative KPS</td>
<td>Median (IQR) 90 (80, 90)</td>
</tr>
<tr>
<td>RTOG RPA, % (n)</td>
<td>Class III 20 (18)</td>
</tr>
<tr>
<td></td>
<td>Class IV 19 (17)</td>
</tr>
<tr>
<td></td>
<td>Class V 45 (41)</td>
</tr>
<tr>
<td></td>
<td>Unable to classify 17 (15)</td>
</tr>
<tr>
<td>Initial treatment, % (n)</td>
<td>XRT 23 (21)</td>
</tr>
<tr>
<td></td>
<td>XRT with concurrent TMZ 77 (70)</td>
</tr>
<tr>
<td>Bevacizumab therapy – % (n)</td>
<td>Yes 25 (23)</td>
</tr>
<tr>
<td></td>
<td>No 75 (68)</td>
</tr>
<tr>
<td>Antiepileptic use, % (n)</td>
<td>Yes 69 (63)</td>
</tr>
<tr>
<td></td>
<td>No 31 (28)</td>
</tr>
<tr>
<td>Corticosteroid use, % (n)</td>
<td>Yes 78 (69)</td>
</tr>
<tr>
<td></td>
<td>No 22 (20)</td>
</tr>
</tbody>
</table>

Abbreviations: IQR, interquartile range; XRT, external-beam RT; TMZ, temozolomide.
of death for the dichotomized variables in these multivariate models are shown in Table 3. Sensitivity analyses demonstrated that the number of impaired tests within CTB Comp was significantly associated with survival (P < .0001).

Kaplan–Meier curves of survival by impairment status on the 4 neurocognitive tests significant at the P ≤ .01 level are shown in Fig. 1. For COWA, median survival was 24.4 months in the unimpaired group and 13.3 months in the impaired group. For Digit Span, median survival in the 2 groups was 23.6 and 11.3 months, respectively. Stratified by TMTB, median survival was 47.7 months and 13.3 months, respectively. Finally, for Similarities, median survival was 23.6 and 9.1 months, respectively. Of the 4 tests in which impairment was significantly associated with survival, 37.8% of patients were impaired on none, 36.7% of patients were impaired on 1 test, 13.3% of patients were impaired on 2 tests, 6.7% were impaired on 3 tests, and 5.6% were impaired on all 4. Of patients impaired on only 1 test, 75.8% were impaired on TMTB, the balance were impaired on COWA. No patients were impaired on only Digit Span or only Similarities. Figure 2 displays a Kaplan–Meier plot of
survival by number of impaired tests amongst the 4
tests associated with survival in the dichotomous
models ($P = .0001$). Patients with impairment on 2,
3, and 4 tests were combined into a single group, due
to lack of statistically significant survival differences
among them.

RTOG RPA classification was possible in 76 patients.
Fifteen patients could not be assigned an RPA class given
lack of information on ability to work. Kaplan–Meier
curves for each of the 4 tests in subgroups defined by
RPA classes III, IV, V, and unclassifiable are displayed
in Fig. 3, with associated log-rank $P$ values. No patients
within RPA class IV were impaired on either Digit Span
or Similarities.

In the more complex model including volumetric
extent of resection, initial treatment, and confirmed
receipt of bevacizumab, COWA remained a significant
predictor of survival both as a continuous ($P = .0063$)
and as a dichotomized ($P = .0032$) variable, and
TMTB approached significance as a continuous variable
($P = .0292$) and was significant when dichotomized ($P = .0017$).

TMTA ($P = .0070$), Digit Span ($P < .0001$), and Similarities ($P = .0050$) dichotomized
by impairment status also retained statistical significance,
while Token approached significance ($P = .0157$). In the subset of patients who received all
3 portions of the HVLT-R, the continuous normalized
HVLT-R DR ($P = .0101$) and HVLT-R Recog
($P = .0251$) scores again approached statistical significance, and the CTB Comp score was significantly

Fig. 1. Kaplan–Meier plot of survival by impairment on (A) COWA, $P = .0029$; (B) Digit Span, $P = .0003$; (C) Similarities, $P < .0001$; and
(D) TMTB, $P < .0001$.

Fig. 2. Kaplan–Meier plots of survival by number of tests in
impaired range amongst COWA, Digit Span, Similarities, and
TMTB.

Fig. 2. Kaplan–Meier plots of survival by number of tests in
impaired range amongst COWA, Digit Span, Similarities, and
TMTB.
associated with survival as a continuous variable ($P = .0089$) with a trend toward significance when dichotomized at the median ($P = .0139$). Of the added clinical variables, initial treatment with radiation plus temozolomide was associated with or showed a trend toward prolonged survival in all models. Neither extent of resection nor bevacizumab use was significant in any model (data not shown).

**Discussion**

In this study, cognitive function was an independent predictor of survival time in patients with newly diagnosed glioblastoma. Specifically, performance on COWA and TMTB was associated with survival when either analyzed as a continuous variable or dichotomized by impairment. Differences in median survival time between impaired and unimpaired patient groups suggest that this finding is clinically significant. Further, these tests predicted patient prognosis within subgroups defined by RTOG RPA class, a widely used and well-validated prognostic scoring system. Impairment on Digit Span or Similarities was also associated with shorter survival but only when dichotomized by impairment, which was never seen without coexisting impairment in either COWA or TMTB.

Patient performance status is a significant predictor of survival in patients with many types of brain tumors,
including high-grade glioma, low-grade glioma, and brain metastases. Performance status is a multifaceted construct, containing elements of both cognitive and motor function. Given that brain tumors can lead to relatively isolated deficits, global measures of patient function may not capture clinically important patient prognostic variation if deficits in cognitive and motor domains have different implications for survival. Previous reports have demonstrated the prognostic significance of performance on neuropsychological test batteries in a mixed sample of patients with recurrent anaplastic astrocytoma and glioblastoma and amongst older patients with newly diagnosed glioblastoma. Other studies have evaluated the association between survival and performance on the Mini-Mental State Exam (MMSE), showing that impairment on the MMSE is associated with shorter survival in patients with low-grade glioma and newly diagnosed glioblastoma. While the MMSE is useful in that it is widely known and easily administered, it does not allow for interrogation of specific cognitive domains, and the commonly used definition of impairment, a score of less than 27 out of 30 possible points, does not account for patient-to-patient variation based on factors such as age and education. To our knowledge, this is the first report documenting the prognostic significance of individual neuropsychological tests specific to the cognitive domain, as well as the CTB Comp, in patients with newly diagnosed glioblastoma, and the first report to demonstrate the ability to stratify prognosis within patient groups defined by RTOG RPA class via cognitive testing. Of note, no association between upper extremity weakness as measured by Grip Strength and survival was seen in this patient group, suggesting that cognitive impairment may be more strongly associated with survival than motor impairment. Ruden et al. recently reported that an objective assessment of ambulatory motor ability as measured by the 6-minute walk test was not associated with survival in high-grade glioma patients.

Based on the pattern of tests found to be predictive of survival, the domains of attention and executive function appear to be particularly relevant to survival in newly diagnosed patients with glioblastoma. The COWA and TMTB tests were significant as both continuous and categorical variables. Notably, they represent 2 out of 3 cognitive tests that are used in most brain tumor clinical trials to date and are amongst the neuropsychological tests recommended by the Response Assessment in Neuro-Oncology and the International Cognition and Cancer Task Force for routine inclusion in studies evaluating cognition in patients with cancer. TMT performance is highly vulnerable to the effects of brain injury, may be impaired in patients with minor head injury or early in the course of dementing disease, and is predictive of both physical impairment and mortality in community-dwelling older adults. In addition to requiring executive function, the COWA task requires intact expressive language, and it is often impaired in patients with expressive aphasia. However, the lack of association between prognosis and Naming, which measures expressive language with greater specificity than does COWA, indicates that the executive function aspects of COWA are the more important ones with respect to survival. This is also consistent with the lack of association between tumor laterality and survival, as executive function is broadly distributed, whereas if language function itself were predictive of survival, patients with left-sided tumors would be anticipated to have poorer prognoses. The CTB Comp was associated with survival as a continuous variable but failed to reach the a priori level of statistical significance as a dichotomized variable. Similarly, there was evidence that aspects of learning and memory (ie, on HVLT-R) were also associated with survival; however, these trends failed to reach our a priori level of statistical significance. The limited sample size available for the learning and memory and the CTB Comp variables likely adversely affected our power to detect significant differences.

Associations between cognitive test results and survival were most frequently seen in patients within RTOG RPA class V. This may simply be due to the smaller patient numbers in other RPA classes limiting our power to detect differences in these groups. Alternatively, it is possible that RPA class V contains more prognostic heterogeneity than other classes, facilitating further subdivision. Given the issue of sample size in these subgroup analyses, we would not conclude that cognition does not predict survival of patients in other RPA classes without further evaluation. Likewise, the rates of impairment varied significantly among different cognitive tests, with implications for detecting statistically significant associations. Impairment was defined uniformly for all tests as a z-score of −1.5 or less, to reflect neuropsychological practice. Future studies using alternative thresholds may increase the discriminatory ability of the tests we have identified or demonstrate that other cognitive tests are associated with survival.

While this study provides strong evidence that cognitive dysfunction is an important predictor of survival in patients with newly diagnosed glioblastoma, further research is necessary to confirm and extend these findings. The patient group evaluated in this study was subject to the referral effect, in which patients with significant functional disabilities or tumors not amenable to extensive resection were less likely to receive cognitive evaluation. This preselection process may explain the lack of significant association of factors such as KPS score and extent of resection with survival in this patient group. While this effect limits the generalizability of these results, it also results in a population enriched with the patient group in whom the predictive power of cognitive testing may be of greatest interest, those with sufficient performance status to be considered candidates for aggressive therapy. Additional research will be needed to determine whether performance in specific cognitive domains is as strongly associated with survival in less selected patient groups. Ideally, such analysis would take place in the setting of a prospective clinical trial, in which information on tumor molecular characteristics (eg, MGMT
methylation, \(IDH\) mutation, gene expression analysis) and treatment at time of tumor progression are systematically recorded. Such a study would allow for construction of a more comprehensive prognostic scoring system in which detailed patient-level and tumor-level characteristics can be evaluated in tandem.

In sum, we demonstrate for the first time the prognostic importance of performance in specific cognitive domains as well as the CTB Comp scores for patients with newly diagnosed glioblastoma. Several individual tests showed statistically significant associations with survival, and each of these tests shared a common dependence on attention and executive functioning, indicating that function in these domains may be especially important for survival. Ongoing prospective trials incorporating longitudinal cognitive testing will be able to evaluate these associations in a broader patient population, such as individuals initially treated with biopsy rather than extensive resection. Similarly, the relationship between tumor molecular markers and cognition and survival can be interrogated with these data. Ultimately, the development of a more robust clinical prognostic classification system, to be used in conjunction with prognostic information based on tumor pathology, will allow for more accurate patient counseling and trial stratification.

Acknowledgments

This material has not been previously published or presented in any venue.

Conflict of interest statement. None declared.

Funding

Portions of the data for this work have been obtained through a search of the integrated multidisciplinary Brain and Spine Center Database. The Brain and Spine Center Database was supported in part, by an institutional MD Anderson database development grant.

References


