Title

Joint Models for the Analysis of Patient Reported Outcome and Neurocognitive Function Data in the Presence of Non-Ignorable Missing Data

Short/Running Title Joint Models for Non-Ignorable Missing Data

Author Names

Stephanie L Pugh[1], Jason Mao[2], Mark R Gilbert[3], Jeffrey S Wefel[4], Terri S Armstrong[3], Stewart Anderson[2,5]

Author Institutions

[1]NRG Oncology Statistics and Data Management Center, American College of Radiology, Philadelphia, PA, USA

[2]Department of Biostatistics, University of Pittsburgh, Graduate School of Public Health, Pittsburgh PA, USA

[3]National Cancer Institute, Bethesda, MD, USA

[4]The University of Texas MD Anderson Cancer Center, Houston TX, USA

[5]NRG Oncology Statistics and Data Management Center, University of Pittsburgh, Pittsburgh PA, USA

First Author's Name

Stephanie L Pugh

50 South 16th Street, Suite 2800

Philadelphia, PA 19102

pughs@nrgoncology.org

215-717-0850

Conflict of Interest Statement

Dr. Anderson has nothing to disclose. Dr. Armstrong reports grants from Genentech, from null, outside the submitted work. Dr. Gilbert has nothing to disclose. Jason Mao has nothing to disclose. Dr. Pugh reports other from Genentech-Roche, grants from Pennsylvania Department of Health, during the conduct of the study; other from Millennium, other from Pfizer-Astellas, outside the submitted work. Dr. Wefel reports personal fees from Roche, during the conduct of the study; personal fees from AbbVie, grants from AngioChem, grants and personal fees from Bayer, personal fees from blueprint medicines, grants from Juno, grants from Novocure, grants and personal fees from Vanquish Oncology, grants from Roche, outside the submitted work.

Funding Statement

This project was supported by grants U10CA180868 (NRG Oncology Operations), U10CA180822 (NRG Oncology SDMC), UG1CA189867 (NRG Oncology NCORP) from the National Cancer Institute (NCI) and Genentech. This project is funded, in part, under a Grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

Data Availability Statement

The data from the present publication will be made available by request from the NCTN/NCORP Data Archive: https://nctn-data-archive.nci.nih.gov.

Keywords: Joint models; non-ignorable missing data; patient reported outcomes

www.ClinicalTrials.gov NRG-RTOG 0825 Identifier: NCT00884741

Joint Models for Non-Ignorable Missing Data

Abstract

Background/Aims

NRG/RTOG 0825 assessed the efficacy of bevacizumab as treatment of newly diagnosed glioblastoma that collected patient reported outcomes (PROs) and neurocognitive tests in progression-free patients. This resulted in non-ignorable missing data since it was more prevalent on the placebo arm due to the bevacizumab arm's numerically increased time to progression or death.

Methods

Joint models were used to assess whether longitudinal PRO and neurocognitive function data predicted time to progression or death. An association test was performed to test the significance of the shared parameter between the two processes. Linear mixed effects models were used for the longitudinal process and time-dependent Weibull relative risk models were used for the time-to-event process. Adjustments were made for recursive partitioning analysis (RPA) class and O(6)-Methylguanine-DNA methyltransferase (MGMT) activity status.

Results

The association tests were significant in nine of the 12 measures assessed. Significant treatment effects were found on the trajectories of many measures of symptom burden (composite symptom score, composite symptom interference, neurologic factor) and measures of quality of life (cognitive functioning, motor dysfunction, and communication deficit) as well as on a single neurocognitive function measure, Controlled Oral Word Association test. For all of these associations, poorer outcomes in patients randomized to receive bevacizumab compared to placebo were observed.

Conclusions

Although there was improvement in PFS in favor of the arm that received bevacizumab that led to non-ignorable missing data imbalanced by arm, significantly worse outcomes were observed for symptom burden, quality of life, and neurocognitive functioning.

Joint Models for Non-Ignorable Missing Data

Citation

Citation

Pugh SL, Moa J, Gilbert MR, Wefel JS, Armstrong TS, Anderson S. Joint models for the analysis of patient-reported outcome and neurocognitive functioning data in the presence of non-ignorable missing data. NRG Oncology Unpublished Research Series. 2022 Sep 2

NRG Oncology Unpublished Research Series. Uploaded 2025 Jun 4

https://www.nrgoncology.org/Portals/0/Publications/Unpublished%20Research%20Series/Pugh%200825%20CURE%20(M%2017124)%20Manuscript.pdf?ver=AqhTY6hE0UvHTEmcKivNQA%3d%3d

Introduction

Patient reported outcomes (PROs) assess quality of life (QOL) and symptoms experienced by the patient directly from the patient. They are typically collected longitudinally and are burdened by not only site compliance and data submission issues but also by non-compliance on the part of the patient. Tests of neurocognitive function, although not reported directly by the patient but rather, are an objective measurement of cognition in which a test administrator works with the patient to complete the tests, can suffer from the same patient non-compliance issues. This missing data can affect the type of analysis that is conducted to reduce the bias missing data introduces.

Briefly, datasets with ignorable missing observations can be analyzed using complete case methods that require listwise deletion of cases with missing data and don't bias the results of the analysis although do reduce power by decreasing the sample size. Missing data that is dependent on other extraneous factors, such as age, socioeconomic status, or grade of tumor, can also be considered ignorable if the analysis method properly adjusts for these factors. Missing data that is dependent on the missing value itself is considered non-ignorable, requiring more complex analysis methods such as shared parameter or joint models.

NRG/RTOG 0825 was a phase III double-blinded placebo-controlled clinical trial assessing the efficacy of bevacizumab in the treatment of newly diagnosed glioblastoma in which patients could consent to participate in a PRO/neurocognitive component, called the Net Clinical Benefit (NCB) substudy. NCB assessments were no longer collected once a patient progressed since the patient moves on to other treatment and the population of interest for the NCB substudy was progression-free patients. Gilbert et al reported an improvement, albeit not statistically significant per study design due to having co-primary endpoints, in progression-free survival

(PFS; hazard ratio [HR]=0.79, 95% confidence interval [CI]=0.66-0.94, p=0.007) [1]. NCB data collected over time was analyzed several ways, including by deterioration status of the relevant NCB scores at each time point, time to neurocognitive, symptom burden, or QOL deterioration, and using linear mixed models that allow adjustment by other factors that may affect the reason for missingness while also allowing inclusion of all available data as opposed to patients with only complete data [1,2]. These models, however, do not appropriately account for data missing due to progression and/or death. In this scenario, the reason missing was dependent on the patient's imaging and clinical evaluation, which was worse in the control arm leading to higher rates of attrition in that arm. Analysis by deterioration status requires complete data, creating potential biases in the results if the missing data are considered non-ignorable. Time to deterioration may be more appropriate since assessments at later time points are not needed once a patient has deteriorated but information from these unused assessments is lost. Herein, a different analysis technique, joint models, will be used to interpret the NCB data within the context of PFS as the reason for missing data.

Methods

Assessments

Patients on NRG/RTOG 0825 were given the opportunity to participate in a Net Clinical Benefits (NCB) substudy that involved completion of 2 PROs, the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 and Brain Neoplasms 20 (QLQ-C30 & BN20) and the M. D. Anderson Symptom Inventory – Brain Tumor Module (MDASI-BT), and a neurocognitive function test battery consisting of 3 tests, Hopkins Verbal Learning Test-Revised (HVLT-R), Trail Making Test (TMT) Parts A and B, and Controlled Oral Word Association (COWA) test. These measures were collected at the time of imaging:

approximately 6, 10, 22, 34, and 46 weeks from randomization. Assessment collection ceased upon progression since neurocognitive functioning, quality of life, and symptom scores were expected to be impacted by progression, additionally, with a higher proportion missing as patients are likely to change institutions for salvage treatment.

The MDASI-BT assesses overall symptom burden which includes symptom severity and interference on daily functioning via patient self-report [3]. It is made up of 22 symptoms and 6 items measuring symptom interference with daily life, with identified symptom and interference factor groupings described by Armstrong et al [3]. A 0-10 scale is used to assess the severity of symptom and interference with 0 = not present/did not interfere and 10 = as bad as you can imagine/interfered completely. Total scores are calculated as the average of the items in that scale. Here, symptom severity and three symptom interference scores (global, activities, and mood) along with 6 factor scores (affective, cognitive, neurologic, treatment, generalized/disease, and gastrointestinal) were of interest.

The EORTC QLQ-C30 assesses the health-related quality of life (QOL) of cancer patients using a 30-item self-report questionnaire that has patients rate the items on a 4-point scale, with 1=not at all to 4=very much [4]. The BN20 is a 20-item self-report using the same rating scale that assesses 11 symptom scales/items such as future uncertainty, visual disorder, motor dysfunction, and communication deficit [5]. It is meant for use among brain cancer patients varying in disease stage and treatment modality (i.e. surgery, chemotherapy, radiotherapy, etc.) and should always be complemented by the QLQ-C30. The QLQ-C30 and BN20 are composed of both multi-item scales and single-item measures. The raw score of multi-scale items is calculated as the mean of component items. This score is then standardized such that all of the scales and single-item measures range in score from 0 to 100. A high score for a functional scale represents a healthy

level of functioning, a high score for the global health status represents a high QOL, and a high score for a symptom scale/item represents a high level of symptomatology/problems. In this study, the global QOL score, 3 functioning domain scores (physical, cognitive, and social), and 2 brain tumor specific item (motor dysfunction and communication deficit) are of interest.

The neurocognitive function component consists of three tests, the HVLT-R, TMT Parts A and B, and COWA test, which are combined to determine the Clinical Trials Battery Composite (CTB COMP) score [6,7]. Scores from these tests are converted into standardized Z-scores that adjusts scores for patient's age and education level [8-10].

The HVLT-R assesses verbal learning and memory. The test involves memorizing a list of 12 words for 3 consecutive trials, recalling the 12 words after a 20-minute delay, and then identifying the 12 words from a list of semantically related or unrelated words. Raw scores are derived for each of the three components.

The COWA is a verbal fluency test that measures spontaneous production of words beginning with designated letters. Patients are given 1 minute to name as many words as possible beginning with the designated letter. The procedure is then repeated for the remaining two letters. The score is the sum of the correct responses with a higher score indicating better functioning.

The TMT is a test requiring visual scanning, psychomotor processing speed, visual attention and task switching. The test can provide information about visual search speed, scanning, speed of processing, mental flexibility, as well as executive functioning. The subject is instructed to connect a set of 25 dots as quickly as possible while still maintaining accuracy. There are two parts to the test: in the first (Part A), the targets are all numbers (1, 2, 3, etc.) and the test taker needs to connect them in sequential order; in the second part (Part B), the subject alternates

between numbers and letters (1, A, 2, B, etc.). The score is the amount of time, in seconds, that it takes the patient to complete each maze such that a lower score indicates better functioning.

Statistical Methods

A subset consisting of 12 of the 23 scores analyzed as part of the NCB component of NRG/RTOG 0825 were selected for these joint model analyses. The 23 scores were from the neurocognitive testing battery (Z-scores from the HVLT-R, TMT Part A and B, and COWA test, as well as the CTB COMP), EORTC QLQ-C30 (global QOL, three functioning domains, and two brain tumor specific items), and the MDASI-BT (symptom severity, three symptom interference, and six factor scores). Scores were selected to have a variety of ranges of the scores, types of scores, and reported significance of the treatment effect. This allows examination of the joint modeling framework in all these scenarios. Table 1 provides a summary of all scores analyzed as part of the NRG/RTOG 0825 NCB component and those analyzed here.

Since the NCB sub-study of NRG/RTOG 0825 only evaluated progression-free alive patients, there are non-ignorable missing data in which the amount of missingness differs between treatment arms. To best account for this in the statistical analysis, a joint model framework was implemented. Joint modeling allows one to characterize the relationship between repeatedly measured data and a time-to-event response. Hence, the modeling process allows researchers to understand how the trajectory of a longitudinal process is related to a time-to-event process. In the case of NRG/RTOG 0825, we wish to relate the longitudinal NCB data and time to progression or death. Following the notation of Rizopoulos [8], let $m_i(t)$ = the *true* trajectory of the longitudinal process for subject *i*; $M_i(t)$ be the *true history* of $m_i(t)$ through time *t* and $\mathbf{u}_i$ be a covariate vector for subject *i*; potentially related to the failure process [11]. Then, the following model is considered:

$$S_i(t|M_i, \mathbf{u}_i) = Pr(T^* > t|M_i(t), \boldsymbol{u}_i) = exp\left(-\int_0^t h(s)exp[\boldsymbol{\gamma}'\boldsymbol{u}_i + \alpha m_i(t)]ds\right).$$

Using a shared parameter, α, between the longitudinal and survival models, the joint distribution can be fully represented that accounts for the non-ignorable missing data [12]. A commonly used structure for the joint model is a random effects joint model that uses latent variables to summarize the associations between the survival and longitudinal outcomes [11,12]. The joint model has the structure of a mixed effects linear model for the longitudinal sub-model and allows for a variety of models for the survival sub-model, such as a time-dependent Weibull relative risk model or a Cox proportional hazards model, with this survival sub-model also containing the effect of the true longitudinal outcome as a linear predictor [11,13].

Mixed-effects linear models, estimated using restricted maximum likelihood (REML), were built with either a random intercept or random intercept and slope, with the best model chosen by comparing BIC. The fixed (population) effects were modeled using linear, quadratic or cubic polynomials, depending on the nature of the trajectory of each longitudinally measured variable. Time (including both quadratic and cubic) and a treatment arm by time interaction were included in longitudinal models if significant at the 0.10 significance level. Time was a continuous variable measured as time of assessment from randomization, with baseline (pre-treatment) as 0 weeks. Depending on the distribution of each outcome variable, transformations, such as the square root, log + a constant, and cubed root, were considered. The initial models performed by Gilbert et al [1] also included higher order time variables but time was treated continuously according to time point of assessment (i.e. 0, 6, 10, 22, 34, and 46 weeks) which was determined by specifying time windows around each time point. No transformations were considered in the initial models previously reported. Covariates included in all models were treatment arm, RPA class and MGMT status. When covariates are omitted from the Cox model, estimates of the

regression parameters may be asymptotically biased towards zero [14]. Time was initially coded in weeks but was changed to years, as needed, if there were convergence issues with the joint model.

PFS was modeled using a time-dependent Weibull relative risk model. The same structure of the longitudinal process as the mixed models was used for the joint models where possible. Wald tests were conducted to determine the effect of treatment arm on the longitudinal process. Adaptive Gauss-Hermite integration method was employed for estimating joint model parameters. A test of the association parameter, α, in the equation above, is a test of the shared parameter that measures the association between the longitudinal metric and the risk for progression. The association point estimate and associated 95% CI and p-value are all of interest as they determine the direction and strength of the effect that the trajectory of each longitudinal variable has on PFS. A negative association value indicates that a positive trajectory is associated with worse PFS on average whereas a positive association value indicates that a positive trajectory is associated with better PFS on average.

Significance for all treatment effects was set at the 0.01 level, which is a more conservative approach due to the issue of multiple comparisons within each of the three components of the NCB: neurocognitive function, QOL, and symptom burden. This practical approach was chosen since a Bonferroni adjustment may be too conservative due to the correlation between each of the measures. Significance for other tests, such as the association test, will be at the 0.05 level. Longitudinal sub-models were fit using SAS Proc Mixed and the joint models were performed using the R package JM [15, 16].

Results

The distribution of each of the outcome measures was examined to assess departure from normality. Only the MDASI-BT scores were problematic due to being highly left skewed (Figures 1a-e). Although none of the considered transformations seemed adequate, a square root was used for the analyses. Histograms for the neurocognitive test and QOL scores are provided in Supplemental Figures 1-7. Full model results are provided in Supplemental Tables 1-12.

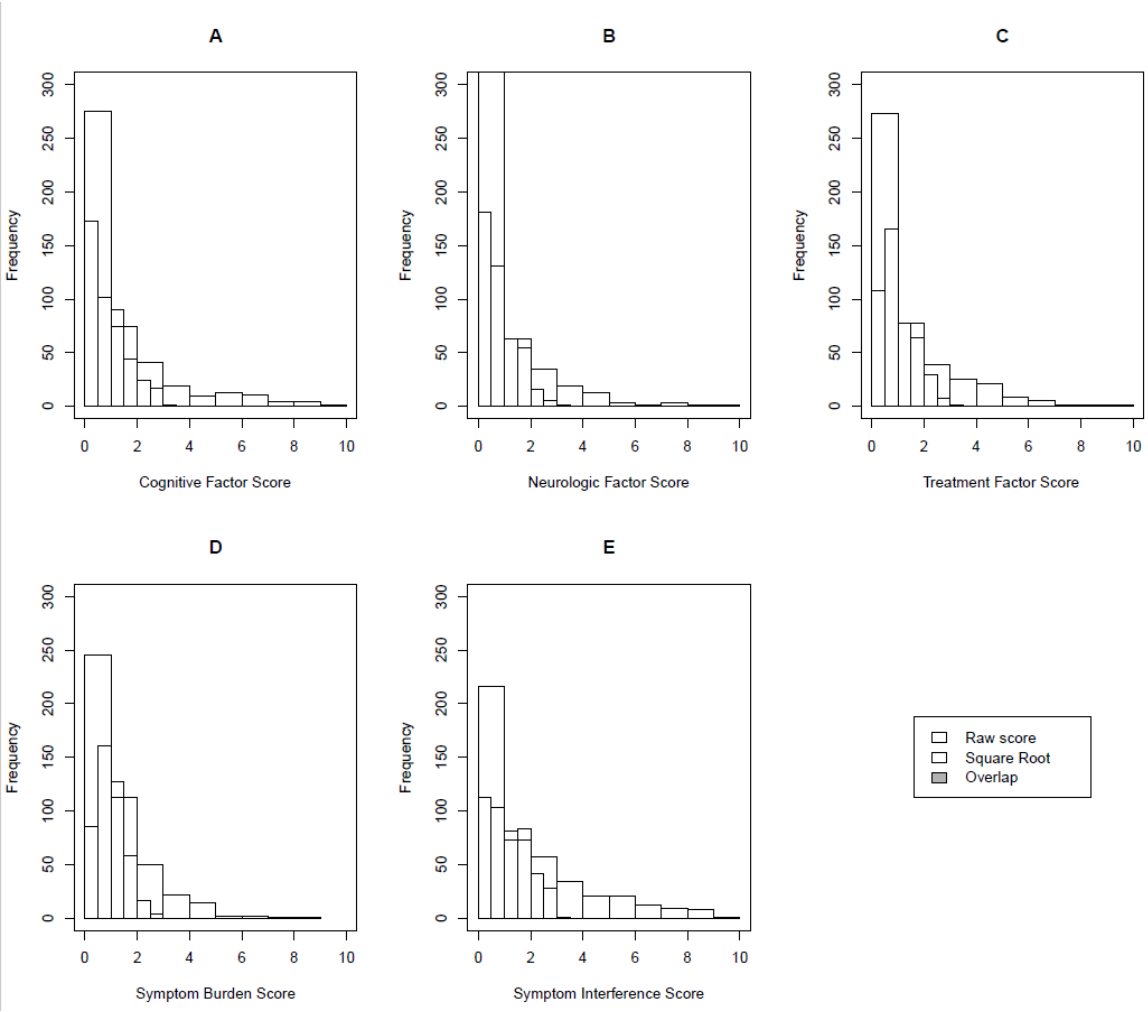


Figure 1- Symptom Baseline Distributions

Neurocognitive Testing

The association tests were significant for all three neurocognitive function outcomes assessed. For TMT Part A score, the biggest impact from the joint model was the significance of time in

which the estimates of time, time^2, and time^3 moved farther from 0 as compared with the mixed model (5.75, -17.23, 12.32, respectively, in the longitudinal sub-model vs. 4.86, -14.25, and 10.31 in the mixed effects model; Supplemental Table 1a-b). The effect of the interaction between treatment and time was not significant in the mixed effects model (estimate = -0.0070; 95% CI: -0.01, -0.00050; p = 0.036) but was significant in the longitudinal sub-model (estimate = -0.20; 95% CI: -0.014, -0.0027; p = 0.0037, Table 2, Supplemental Table 2a-b). For the composite score, the only difference between the two models was for RPA class IV (as compared to class III; Supplemental Table 3a-b). The effect of treatment remained insignificant in both models.

Symptom Assessment

All of the MDASI-BT scores were transformed using a square root. Results are presented on the square root scale. The association tests for cognitive factor and composite symptom score were not significant and thus, there were no notable differences between estimates in either model (Table 3, Supplemental Tables 4a-b and 5a-b). For neurologic factor, the effect of treatment became significant in the joint model (estimate = -0.16; 95% CI:-0.28, 0.043; p = 0.0075) as it was not in the mixed effects model (estimate = -0.14; 95% CI:-0.25, -0.020; p = 0.022; Table 3, Supplemental Table 6a-b). For treatment factor, although the effect of treatment remained not significant in both models, the interaction effect had less of an impact in the joint model as compared to the mixed effect model (estimate = 0.30; 95% CI: 0.048, 0.56; p=0.020 vs. estimate = 1.00; 95% CI: 0.33, 1.67; p = 0.0039, respectively). The effects of time and time^2 changed between being non-significant in the mixed effects model (estimate = 1.05; 95% CI: 0.37, 1.73; p = 0.0023 and -4.35; 95% CI:-6.30, -2.39; p < 0.0001, respectively) to significant in the longitudinal sub-model (estimate = -0.031; 95% CI: -0.53, 0.47; p=0.90 and -1.37, 95% CI: -

2.84, 0.10, p=0.068, respectively) for composite symptom interference score (Table 3, Supplemental Table 12a-b).

Quality of Life

The association test was significant for global QOL, cognitive functioning, and motor dysfunction, but not communication deficit. Although significant in both models for global QOL score, the effect of time, time^2, and time^3 only increased in the longitudinal sub-model as compared to the mixed effects model (estimates of -52.14, 164.80, and -116.42, respectively, for the longitudinal sub-model, vs. -41.23, 133.55, and -95.13, respectively, for the mixed effects model; Table 4, Supplemental Table 9a-b). The only notable difference in the models for cognitive functioning was in the treatment-by-time interaction. The joint model p-value, unlike that for the mixed effects model, crossed the significance level threshold of 0.01 (estimate = -0.23; 95% CI: -0.37, -0.078; p = 0.0027 for the joint model vs estimate = -0.19; 95% CI: -0.34, -0.044; p = 0.011 for the mixed effects model; Table 4, Supplemental Table 10a-b). For motor dysfunction, the effects of both treatment and the treatment-by-time interaction increased from the mixed effects model to the joint model with the interaction term becoming significant (treatment and interaction estimates -3.23 and 0.15 for the joint model vs. -2.71 and 0.11 for the mixed model; Table 4, Supplemental Table 11a-b). The association test was not significant for communication deficit, however the effect of treatment was not significant in the mixed effects model (estimate = -3.95, 95% CI: -7.42, -0.49; p = 0.026; Table 4, Supplemental Table 12a) but was in the longitudinal sub-model of the joint model (estimate = -3.92, 95% CI: -6.84, -0.99; p = 0.0086; Table 4, Supplemental Table 12b). In the longitudinal sub-model, the estimate of the effect actually decreased and the confidence interval became tighter.

Discussion

In NRG/RTOG 0825, 45.2% of patients randomized to the control arm were thought to have disease progression or died by 6 months compared to only 19.3% of patients randomized to the addition of bevacizumab [1]. While some of these patients with presumed early progression likely had pseudoprogression, which is much less common with bevacizumab use, this contributed to a large amount of missing data, unbalanced by arm, at the 22-, 34-, and 46-week time points in the NCB analysis [17]. In the subset of the 12 NCB endpoints examined in this analysis, eight had a significant test of the shared parameter between the longitudinal and time-to-event processes, indicating that the two are related. This justifies the need to perform joint modeling to account for the non-ignorable missing data.

In this analysis, three of the seven neurocognitive scores, four of the six QOL scores from the EORTC QLQ-C30 and BN20, and five of 10 symptom burden scores from the MDASI-BT were analyzed. The joint models were generally able to be run without convergence errors. Transformations were required for the scores analyzed from the MDASI-BT due to the data, which is scored on a 0-10 scale, being skewed. Convergence issues were resolved by changing how time was coded (weeks vs. years). The significance of the covariates, including time, treatment arm, and the treatment-by-time interaction did change from the mixed model to the joint model, but the overall conclusions of the study remained the same: treatment with bevacizumab generally increased symptom burden, decreased QOL, and decreased neurocognitive function.

In the analysis of NRG/RTOG 0825, general linear models were used with fixed effects for treatment arm, time, MGMT status (methylated vs. unmethylated), RPA class (III vs. IV vs. V), and a treatment by time interaction to assess between-group differences, with a type I error of 0.05 to indicate statistical significance [1]. Neurocognitive function results showed that patients

receiving bevacizumab experienced greater deterioration as compared to patients receiving placebo with respect to the CTB COMP score on the neurocognitive function test battery ($p=0.05$), the COWA test ($p=0.003$), and the TMT Part A ($p=0.04$). Symptom burden results from the MDASI-BT showed that patients receiving bevacizumab experienced greater deterioration than patients receiving placebo with significant between-group differences in the composite symptom score ($p=0.02$), composite symptom-interference score ($p<0.001$), activity-related symptom interference ($p<0.001$), affective factors ($p=0.04$), cognitive factors ($p=0.01$), and treatment factors ($p=0.01$). QOL results from the EORTC QLQ-C30/BN20 showed greater deterioration on the bevacizumab arm than the placebo arm for cognitive functioning score ($p=0.008$), motor dysfunction ($p=0.003$), and communication deficit ($p=0.003$).

Wefel et al [2] additionally reported further NCB endpoints from NRG/RTOG 0825. The focus of this analysis was on deterioration, defined using the reliable change index for the NCF tests and the minimally important difference for MDASI-BT and EORTC QLQ-C30, defined as 1 point and 10 points, respectively [3,7,18-20]. The two central between-arm comparisons were proportion of deteriorated patients for each NCF, QOL, and symptom score at each follow-up timepoint and time to deterioration for NCF, symptom burden, and QOL, defined as the time from randomization to first deterioration on any of the NCF tests, on any MDASI-BT score, and on any EORTC QLQ-C30/BN20 score, respectively. Death and progression were considered competing risks. The results for deterioration status showed mixed results for early time points (6 and 10 weeks) but for later time points (22, 34, and 46 weeks), patients treated with bevacizumab evidenced greater worsening on several tests and scores. No significant differences were seen in time to NCF, symptom burden, or QOL deterioration, with both arms experiencing early deterioration.

The joint modeling analysis showed that patients receiving bevacizumab experienced greater deterioration as a function of time as compared to patients receiving placebo with respect to COWA test (p=0.0037), MDASI-BT composite symptom score (p=0.0077), MDASI-BT composite symptom-interference score (p=0.0003), EORTC QLQ-C30/BN20 cognitive functioning (p=0.0027), EORTC QLQ-C30/BN20 motor dysfunction (p=0.0064), and EORTC QLQ-C30/BN20 communication deficit (p=0.0010). Within the subset of scores without a significant treatment-by-time interaction, patients receiving bevacizumab experienced greater deterioration as compared to patients receiving placebo with respect to MDASI-BT neurologic factor (p=0.0075). It should be noted that the type I error to determine significance is stricter in this analysis than was used by Gilbert et al [1]. Using a significance level of 0.05 would make MDASI-BT cognitive factor (p=0.022) and MDASI-BT treatment factor (p=0.020) also significant.

Longitudinal models in general have several advantages over testing differences at landmark time points, such as the deterioration status analysis, which involved tests conducted at multiple timepoints. These models include adjustment of covariates and use of maximum likelihood estimation to improve estimation in the presence of missing data as well as reducing the inflation of the type I error due to conducting tests at each time point [21,22]. For NRG/RTOG 0825, the time to deterioration was not sensitive to between arm differences due to both arms experiencing a high rate of early deterioration when differences became further pronounced at later time points. The joint model specifically allows estimation of the treatment effect while accounting for the imbalance in completed assessments between arms. For certain outcomes, polynomials were considered for time due to non-linearity. Other options to handle non-linear trajectories are splines [23-25].

Similarly, other options exist to handle the survival submodel, such as parametric models or accelerated failure time models [26]. The Cox proportional hazards model was chosen due to its common use, ease of interpretation, and robustness to misspecification of the underlying survival distribution [27]. A limitation of this analysis is the impact of an incorrectly chosen submodel can have on the results. Ensuring that necessary covariates are used in models and appropriate methods are chosen to handle skewedness of the outcome data and non-linearity in the longitudinal submodel.

In conclusion, the analyses presented here and by Gilbert et al [1] and Wefel et al [2] consistently show a negative impact of bevacizumab on neurocognitive functioning, symptom burden, and quality of life in these newly diagnosed GBM patients. This analysis showed an association between the longitudinal and survival processes suggesting that the analysis of non-ignorable missing data that is driven by a time-to-event process should adjust for the shared parameter between the longitudinal and time-to-event processes to obtain the most precise estimates of the model coefficients. Ignoring the time-to-event process when analyzing the longitudinal data could result in biased results due to the presence of non-ignorable missing data.

References

1. Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, Colman H, Chakravarti A, Pugh S, Won M, Jeraj R, Brown PD, Jaeckle KA, Schiff D, Stieber VW, Brachman DG, Werner-Wasik M, Tremont-Lukats IW, Sulman EP, Aldape KD, Curran WJ, Mehta MP. Randomized Clinical Trial of Bevacizumab for Newly Disagnosed Glioblastoma. NEJM 2014;370(8):699-708. doi: 10.1056/NEJMoa1308573.
2. Wefel JS, Armstrong TS, Pugh SL, Gilbert MR, Wendland MM, Brachman DG, Roof KS, Brown PD, Crocker IR, Robins HI, Lee RJ, Won M, Mehta MP. Neurocognitive, Symptom, and Health- Related Quality of Life Outcomes of a Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma (NRG Oncology RTOG 0825). Neuro-Onc. 2021:noab011. doi: 10.1093/neuonc/noab011.
3. Armstrong TS, Mendoza T, Gning I, Coco C, Cohen MZ, Eriksen L, Hsu MA, Gilbert MR, Cleeland C. Validation of the M.D. Anderson Symptom Inventory Brain Tumor Module (MDASI-BT). J Neurooncol. 2006 Oct;80(1):27-35. doi: 10.1007/s11060-006-9135-z. Epub 2006 Apr 6. Erratum in: J Neurooncol. 2006;80(1):37..
4. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol. 1997;50(4):441-50. doi: 10.1016/s0895-4356(96)00428-3..
5. Osoba D, Aaronson NK, Muller M, Sneeuw K, Hsu MA, Yung WK, Brada M, Newlands E. The development and psychometric validation of a brain cancer quality-of-life

questionnaire for use in combination with general cancer-specific questionnaires. Qual Life Res. 1996;5(1):139-50. doi: 10.1007/BF00435979.

6. Meyers CA, Smith JA, Bezjak A, Mehta MP, Liebmann J, Illidge T, Kunkler I, Caudrelier JM, Eisenberg PD, Meerwaldt J, Siemers R, Carrie C, Gaspar LE, Curran W, Phan SC, Miller RA, Renschler MF. Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. J Clin Oncol. 2004 Jan 1;22(1):157-65. doi: 10.1200/JCO.2004.05.128.
7. Wefel JS, Cloughesy T, Zazzali JL, Zheng M, Prados M, Wen PY, Mikkelsen T, Schiff D, Abrey LE, Yung WK, Paleologos N, Nicholas MK, Jensen R, Vredenburgh J, Das A, Friedman HS. Neurocognitive function in patients with recurrent glioblastoma treated with bevacizumab. Neuro Oncol. 2011 Jun;13(6):660-8. doi: 10.1093/neuonc/nor024.
8. Benedict RH, Schretten D, Groninger L, Brandt J. Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. Clin Neuropsychol. 12, 1998.
9. Ruff RM, Light RH, Parker SB, Levin HS. Benton Controlled Oral Word Association Test: Reliability and updated norms. Arch Clin Neuropsychol. 11:329-338, 1996.
10. Tombaugh TN. Trail Making Test A and B. Normative data stratified by age and education. Arch Clin Neuropsychol. 19:203-214, 2004.
11. Rizopoulos D. Joint models for longitudinal and time-to-event data with applications in R. 1st ed. Boca Raton FL: CRC Press; 2012.
12. Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica 2004;14:809-834.

13. Cox DR. Regression models and life-tables. J Royal Stat Soc (series B) 1972;34:187-229.
14. Gail MH, Wieand C, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1984;71(3):431-444.
15. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. Journal of Statistical Software 2010: 35(9):1-33.
16. Rizopoulos D. Package 'JM'. CRAN respository. April 18, 2018.
17. Zikou A, Sioka C, Alexiou GA, Fotopoulos A, Voulgaris S, Argyropoulou MI. Radiation Necrosis, Pseudoprogression, Pseudoresponse, and Tumor Recurrence: Imaging Challenges for the Evaluation of Treated Gliomas. Contrast Media Mol Imaging. 2018;2018:6828396. doi:10.1155/2018/6828396.
18. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12-9. doi: 10.1037//0022-006x.59.1.12.
19. Maringwa J, Quinten C, King M, Ringash J, Osoba D, Coens C, Martinelli F, Reeve BB, Gotay C, Greimel E, Flechtner H, Cleeland CS, Schmucker-Von Koch J, Weis J, Van Den Bent MJ, Stupp R, Taphoorn MJ, Bottomley A; EORTC PROBE Project and Brain Cancer Group. Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. Ann Oncol. 2011;22(9):2107-2112. doi: 10.1093/annonc/mdq726.
20. Armstrong TS, Wefel JS, Wang M, Gilbert MR, Won M, Bottomley A, Mendoza TR, Coens C, Werner-Wasik M, Brachman DG, Choucair AK, Mehta M. Net clinical benefit

analysis of radiation therapy oncology group 0525: a phase III trial comparing conventional adjuvant temozolomide with dose-intensive temozolomide in patients with newly diagnosed glioblastoma. J Clin Oncol. 2013 10;31(32):4076-84. doi: 10.1200/JCO.2013.49.6067.

21. Gibbons RD, Hedeker D, DuToit S. Advances in analysis of longitudinal data. Annu Rev Clin Psychol. 2010 ; 6:79–107. doi:10.1146/annurev.clinpsy.032408.153550.
22. Brown H, Prescott R. Applied Mixed Models in Medicine. 2nd ed. West Sussex England: John Wiley & Sons Ltd.; 2006.
23. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics. 2011;67(3):819-29. doi: 10.1111/j.1541-0420.2010.01546.x.
24. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1(4):465-80. doi: 10.1093/biostatistics/1.4.465.
25. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics. 2005;61(1):64-73. doi: 10.1111/j.0006-341X.2005.030929.x.
26. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. Stat Med. 2013;32(23):4118-34. doi: 10.1002/sim.5823.
27. Nardi, A, and Schemper, Michael. Comparing Cox and parametric models in clinical studies. Stat Med. 2003;22(23):3597-3610. doi: 10.1002/sim.1592.

Table 1
Summary of NCB Measures

NCB Component	NCB Measure	Range	Treatment-by-Time Interaction p-value	Included in analysis
Neurocognitive functioning	HVLT-R Total Recall	0 to 36	0.613	No
	HVLT-R Delayed Recall	0 to 12	0.390	No
	HVLT-R Delayed Recognition	-12 to 12	0.735	No
	TMT Part A	0 to 2750	0.043[a]	Yes
	TMT Part B	0 to 3750	0.097	No
	COWA	0 to ∞	0.003 [a]	Yes
	CTB COMP	-∞ to ∞	0.050 [a]	Yes
MDASI-BT	Composite Symptom	0 to 10	0.017 [a]	Yes
	Composite Interference	0 to 10	<0.001 [a]	Yes
	Activity-Related Symptom Interference	0 to 10	0.004 [a]	No
	Mood-Related Symptom Interference	0 to 10	<0.001 [a]	No
	Affective Factor	0 to 10	0.038 [a]	No
	Cognitive Factor	0 to 10	0.014 [a]	Yes
	Neurologic Factor	0 to 10	0.135	Yes
	Treatment Factor	0 to 10	0.029 [a]	Yes
	Generalized/disease Factor	0 to 10	0.011 [a]	No
	Gastrointestinal Factor	0 to 10	0.889	No
EORTC QLQ-C30 and BN20	Global QOL	0 to 100	0.048	Yes
	Physical Functioning	0 to 100	0.464	No
	Cognitive Functioning	0 to 100	0.008 [a]	Yes
	Social Functioning	0 to 100	0.101	No
	Motor Dysfunction	0 to 100	0.015 [a]	Yes
	Communication Deficit	0 to 100	0.003 [a]	Yes

NCB=Net Clinical Benefits; HVLT-R=Hopkins Verbal Learning Test-Revised; TMT=Trail Making Test; COWA=Controlled Oral Word Association; CTB COMP= Clinical Trial Battery Composite; MDASI-BT=M. D. Anderson Symptom Inventory-Brain Tumor; EORTC= European Organisation of Research and Treatment of Cancer; QLQ-C30=Quality of Life-Core 30; BN20=Brain Neoplasms 20; QOL=Quality of life.
[a] As reported by Gilbert et al, 2014

Table 2
Comparison of Model Results for Neurocognitive Test Scores

	Treatment Arm			Treatment*Time [a] Interaction			Association Test
	Estimate	95% CI	p-value	Estimate	95% CI	p-value	p-value
TMT Part B							
Longitudinal Model	-0.076	-0.75, 0.60	0.83	N/A	N/A	N/A	N/A
Joint Model	0.024	-0.62, 0.63	0.99	N/A	N/A	N/A	<0.0001
COWA							
Longitudinal Model	-0.076	-0.29, 0.14	0.49	-0.0070	-0.014, -0.00050	0.036	N/A
Joint Model	-0.20	-0.39, -0.011	0.039	-0.0085	-0.014, -0.0027	0.0037	<0.0001
CTB COMP Score							
Longitudinal Model	0.16	-0.16, 0.47	0.33	N/A	N/A	N/A	N/A
Joint Model	0.12	-0.22, 0.47	0.49	N/A	N/A	N/A	0.032

CI=Confidence interval; TMT=Trail Making Test; COWA=Controlled Oral Word Association; CTB COMP= Clinical Trial Battery Composite.
[a] Time was in weeks for COWA test and CTB COMP score and in years for TMT Part A.

Table 3
Comparison of Model Results for Symptom Assessment Scores
MDASI-BT

	Treatment Arm			Treatment*Time [a] Interaction			Association Test
	Estimate	95% CI	p-value	Estimate	95% CI	p-value	p-value
Cognitive Factor							
Longitudinal Model	-0.088	-0.21, 0.034	0.16	0.0048	0.00040, 0.0093	0.034	N/A
Joint Model	-0.097	-0.23, 0.025	0.12	0.0053	0.00078, 0.0098	0.022	0.17
Neurologic Factor							
Longitudinal Model	-0.14	-0.25, -0.20	0.022	0.0048	0.00020, 0.0095	0.042	N/A
Joint Model	-0.16	-0.28, -0.043	0.0075	-0.0077	0.0084, 0.020	0.90	0.0067
Treatment Factor							
Longitudinal Model	-0.077	-0.35, 0.20	0.58	1.00	0.33, 1.67	0.0039	N/A
Joint Model	-0.0097	-0.12, 0.096	0.86	0.30	0.048, 0.56	0.020	0.021
Symptom Burden							
Longitudinal Model	-0.061	-0.35, 0.048	0.14	0.25	0.081, 0.42	0.0040	N/A
Joint Model	-0.059	-0.14, 0.023	0.16	0.23	0.062, 0.40	0.0077	0.50
Symptom Interference							
Longitudinal Model	-0.021	-0.14, 0.021	0.14	0.43	0.17, 0.70	0.74	N/A
Joint Model	-0.046	-0.17, 0.079	0.47	0.50	0.23, 0.48	0.0003	<0.0001

CI=Confidence interval; MDASI-BT=M.D. Anderson Symptom Inventory-Brain Tumor Module.
[a] Time was in weeks for cognitive factor and neurologic factor and in years for treatment factor, symptom burden, and symptom interference.

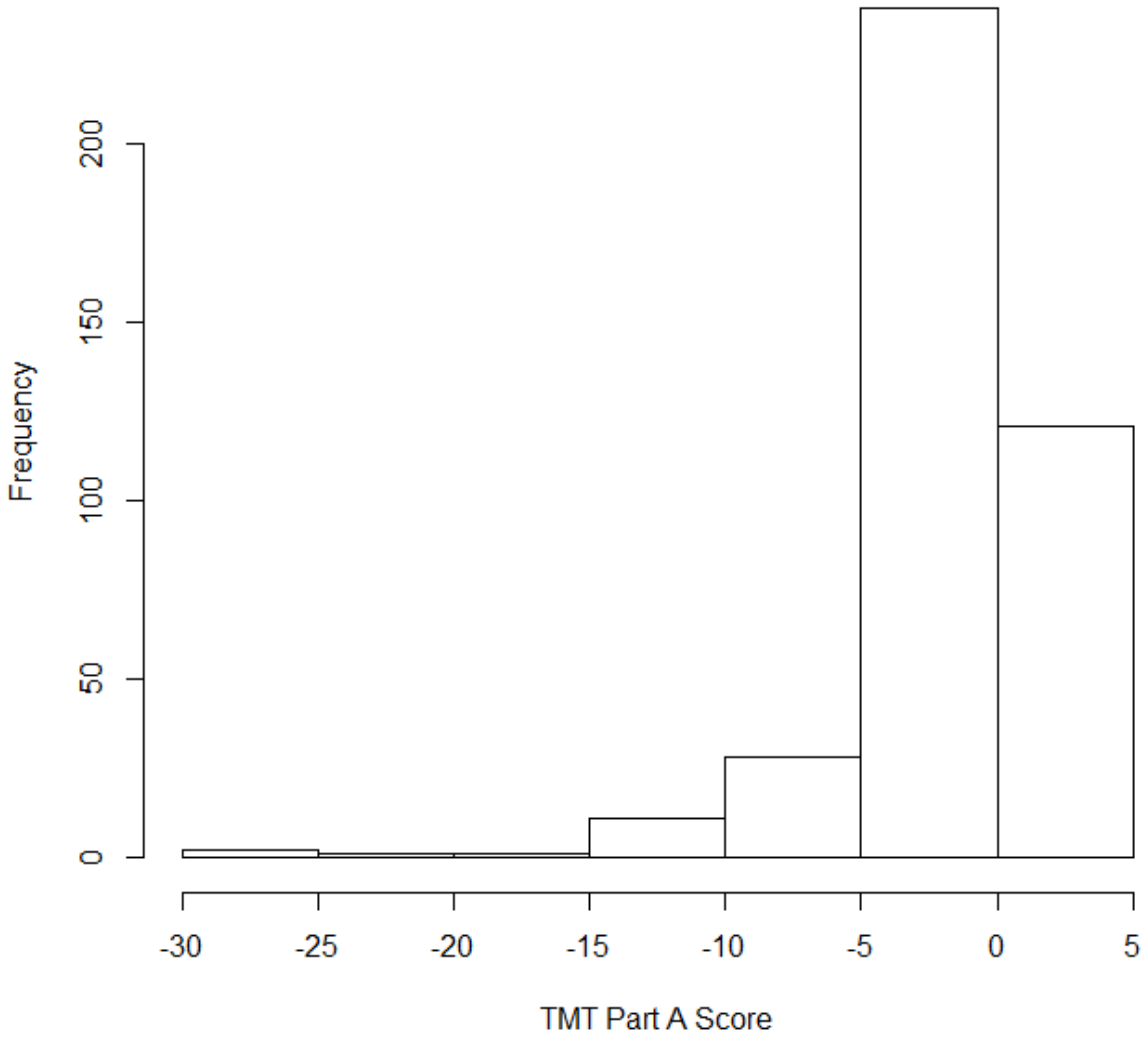
Table 4
Comparison of Model Results for Quality of Life (QOL) Assessment Scores EORTC-QLQ-C30 and BN20

	Treatment Arm			Treatment*Time [a] Interaction			Association Test
	Estimate	95% CI	p-value	Estimate	95% CI	p-value	p-value
Global QOL							
Longitudinal Model	-1.21	-4.34, 1.92	0.45	N/A	N/A	N/A	N/A
Joint Model	-1.26	-4.32, 1.79	0.42	N/A	N/A	N/A	0.024
Cognitive Functioning							
Longitudinal Model	1.42	-2.38, 5.23	0.46	-0.19	-0.34, -0.044	0.011	N/A
Joint Model	1.83	-2.09, 5.74	0.36	-0.23	-0.37, -0.078	0.0027	0.011
Motor Dysfunction							
Longitudinal Model	-2.71	-5.60, 0.19	0.067	0.11	-0.0011, 0.21	0.053	N/A
Joint Model	-3.23	-6.10, -0.35	0.028	0.15	0.042, 0.26	0.0064	<0.0001
Communication Deficit							
Longitudinal Model	-3.95	-7.42, -0.49	0.026	8.42	2.91, 13.93	0.0030	N/A
Joint Model	-3.92	-6.84, -0.99	0.0086	9.22	3.73, 14.71	0.0011	0.14

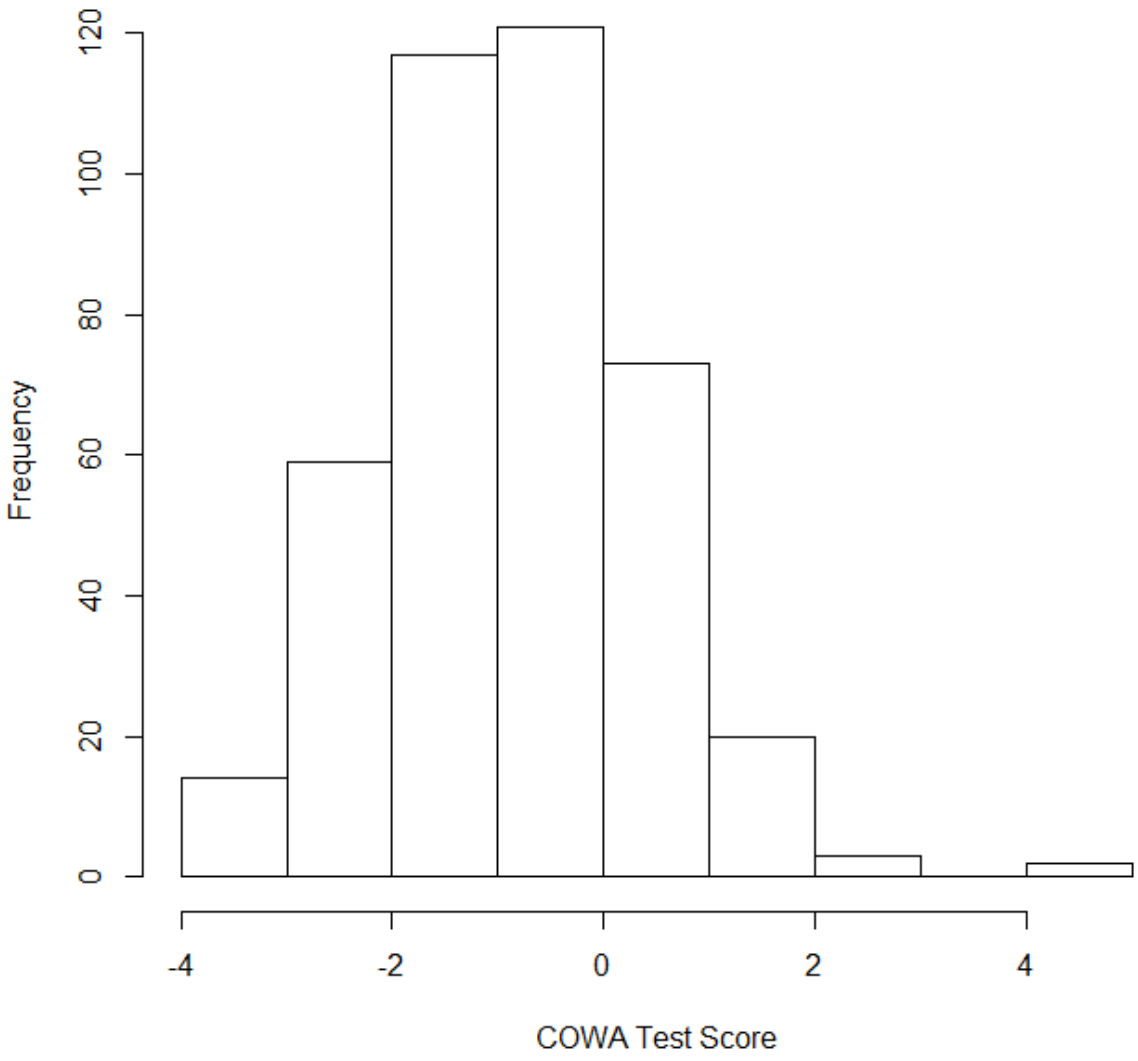
CI=Confidence interval; EORTC=European Organisation of Research and Treatment of Cancer; QLQ-C30=Quality of life questionnaire-Core 30 items; BN20=Brain neoplasms 20 items
[a] Time was in weeks for cognitive functioning and motor dysfunction and in years for global QOL and communication deficit.

Supplemental Materials

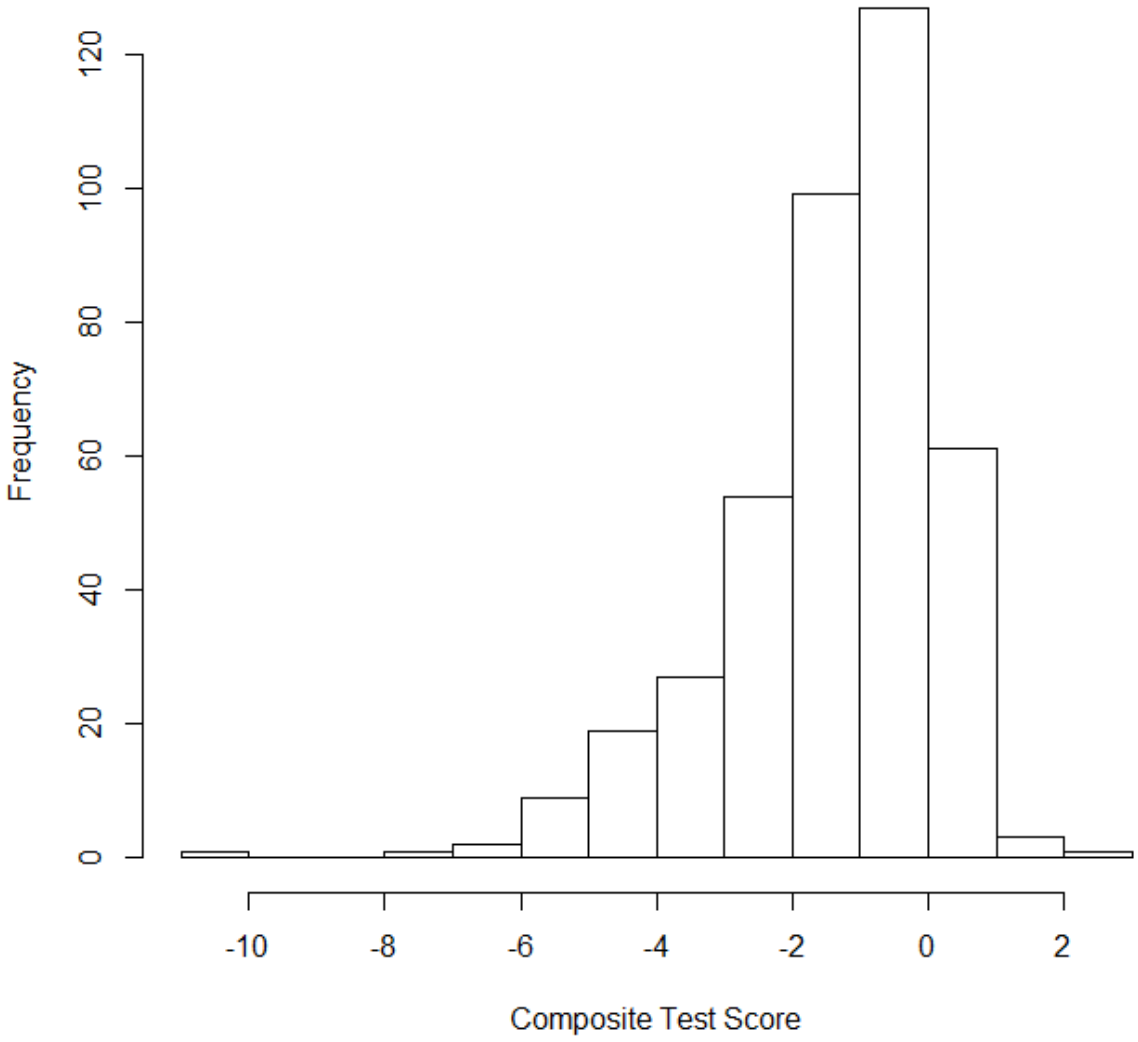
Supplemental Figure 1
Baseline Distribution of Trail Making Test (TMT) Part A Z-Score



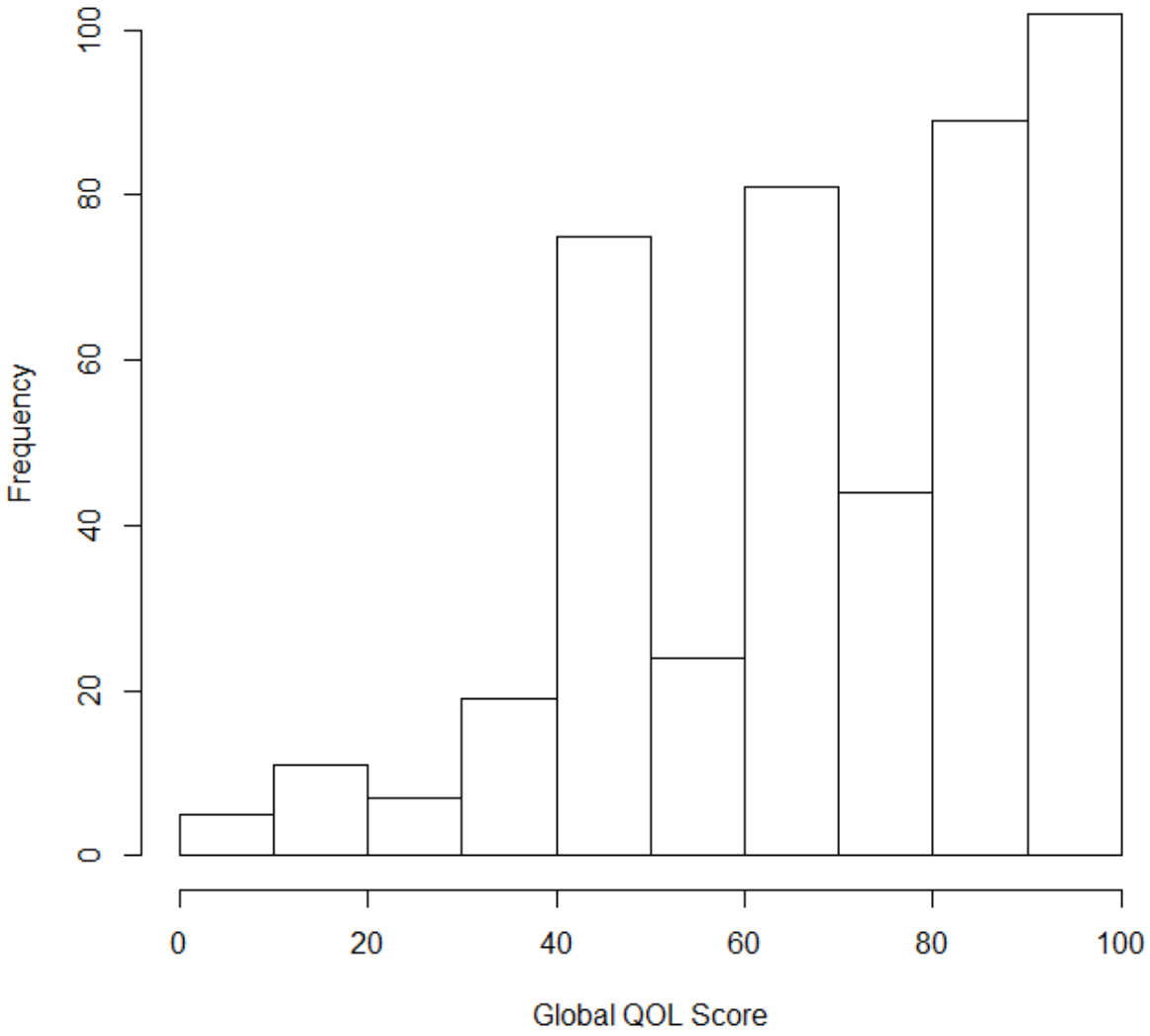
Supplemental Figure 2
Baseline Distribution of Controlled Oral Word Association (COWA) Test Z-Score



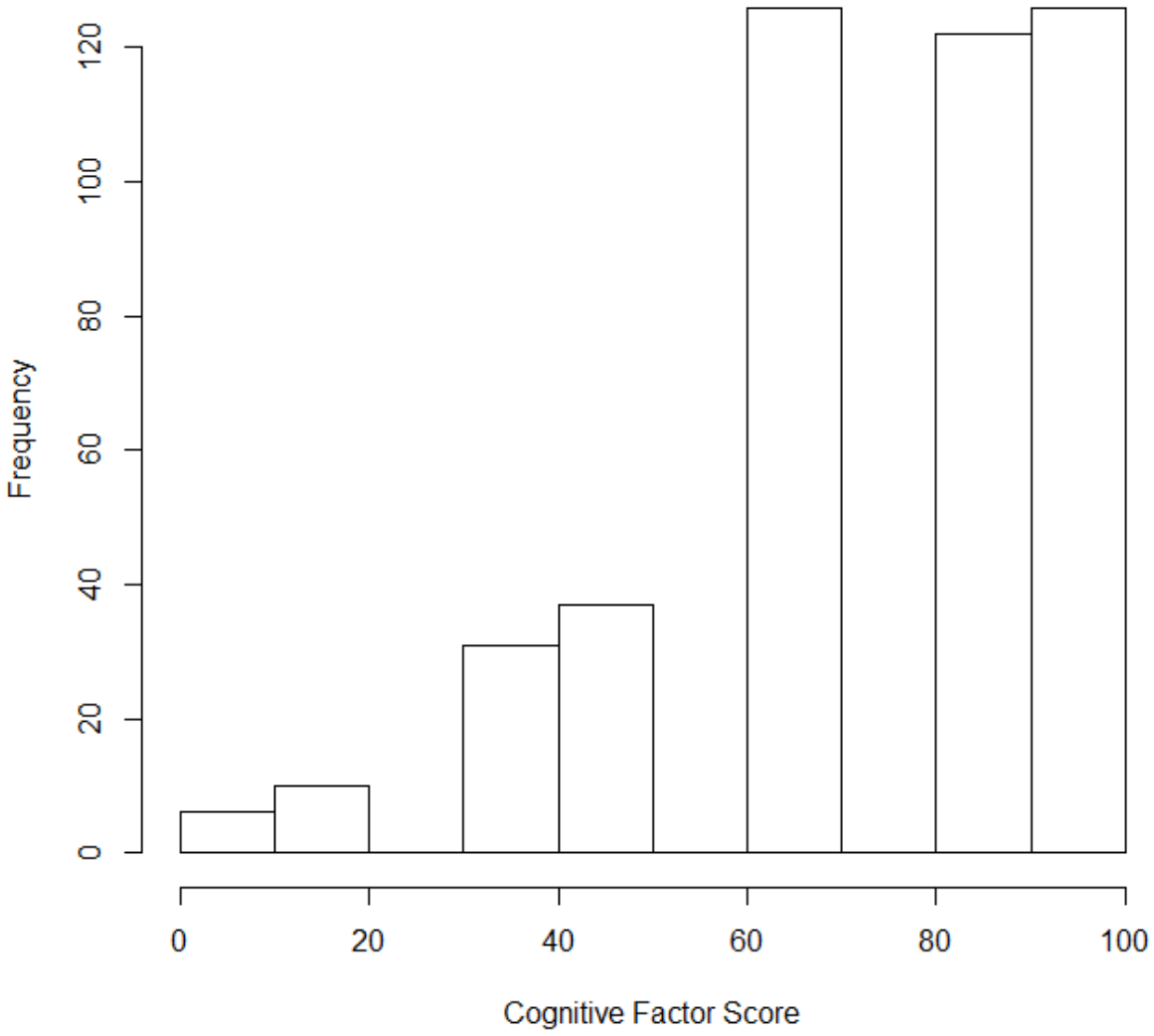
Supplemental Figure 3
Baseline Distribution of the Neurocognitive Battery Composite Score



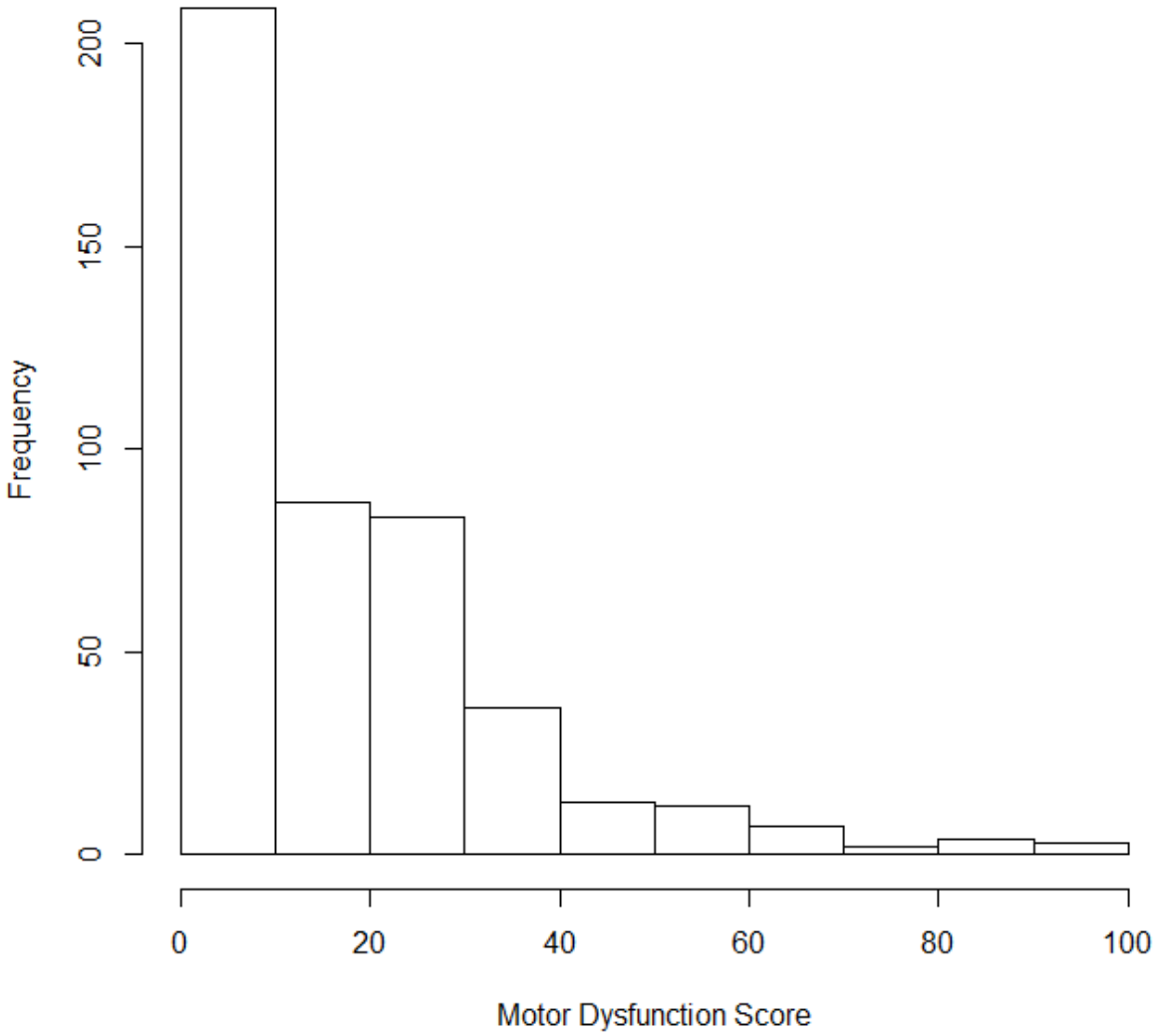
Supplemental Figure 4
Baseline Distribution of Global Quality of Life (QOL) Score



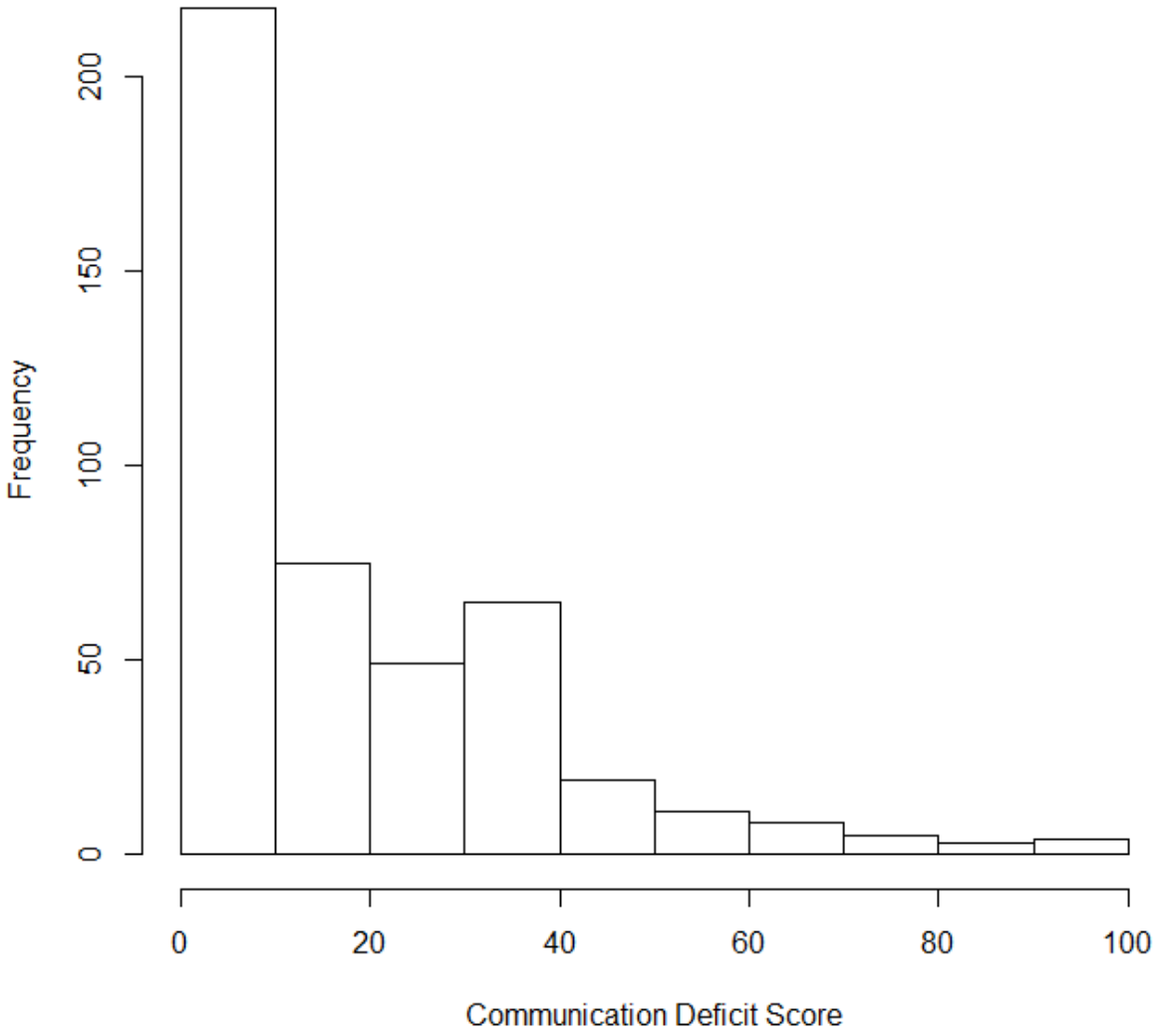
Supplemental Figure 5
Baseline Distribution of Cognitive Functioning Score



Supplemental Figure 6
Baseline Distribution of Motor Dysfunction Score



Supplemental Figure 7
Baseline Distribution of Communication Deficit Score



Supplemental Table 1
Trail Making Test (TMT) Part A Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	-1.45	0.56	-2.56, -0.34	0.011
Time (Years)	4.86	2.08	0.78, 8.93	0.020
Treatment (Arm 1)	-0.076	0.35	-0.75, 0.60	0.83
$Time^2$	-14.25	6.01	-26.05, -2.46	0.018
$Time^3$	10.31	4.52	1.45, 19.71	0.023
MGMT (Methylated)	0.61	0.38	-0.14, 1.36	0.11
RPA IV (RPA III)	-0.54	0.48	-1.49, 0.41	0.26
RPA V (RPA III)	-3.15	0.63	-4.39, -1.90	<0.0001
B. Joint Model				
Longitudinal Sub-Model				
Intercept	-1.49	0.52	-2.50, -0.48	0.0039
Time (Years)	5.75	1.55	2.71, 0.80	0.0002
Treatment arm (Arm 1)	0.024	0.32	-0.62, 0.63	0.99
$Time^2$	-17.23	4.68	-26.40, -8.06	0.0002
$Time^3$	12.32	3.76	4.96, 19.69	0.0010
MGMT (Methylated)	0.55	0.35	-0.14, 1.23	0.12
RPA IV (RPA III)	-0.54	0.44	-1.41, 0.32	0.23
RPA V (RPA III)	-3.28	0.58	-4.43, -2.14	<0.0001
Survival Sub-Model				
Intercept	-0.73	0.22	-1.16, -0.30	0.0010
Treatment arm (Arm 1)	-0.22	0.11	-0.44, -0.0076	0.042
RPA IV (RPA III)	0.71	0.13	0.46, 0.96	<0.0001
RPA V (RPA III)	0.44	0.16	0.12, 0.76	0.0073
MGMT (Methylated)	0.58	0.21	0.16, 0.99	0.0065
Association Test	-0.11	0.022	-0.15, -0.063	<0.0001
log(shape)	0.53	0.069		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept is a random effect.

Supplemental Table 2
Controlled Oral Word Association (COWA) Test Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	-0.49	0.17	-0.83, -0.15	0.0050
Time (Weeks)	-0.076	0.11	-0.29, 0.14	0.49
Treatment (Arm 1)	0.0073	0.0025	0.0024, 0.012	0.0036
MGMT (Methylated)	-0.0070	0.0033	-0.014, -0.00050	0.036
RPA IV (RPA III)	0.15	0.12	-0.087, 0.38	0.22
RPA V (RPA III)	-0.39	0.15	-0.69, -0.094	0.010
B. Joint Model				
Longitudinal Sub-Model				
Intercept	-0.53	0.14	-0.81, -0.26	0.0002
Time (Weeks)	-0.20	0.098	-0.39, -0.011	0.039
Treatment arm (Arm 1)	0.033	0.011	0.012, 0.053	0.0018
MGMT (Methylated)	-0.0085	0.0029	-0.014, -0.0027	0.0037
RPA IV (RPA III)	0.28	0.12	0.047, 0.52	0.018
RPA V (RPA III)	-0.38	0.098	-0.58, -0.19	0.0001
Survival Sub-Model				
Intercept	-6.92	0.39	-7.68, -6.16	<0.0001
Treatment arm (Arm 1)	-0.18	0.11	-0.38, 0.033	0.099
RPA IV (RPA III)	0.66	0.12	0.42, 0.90	<0.0001
RPA V (RPA III)	0.45	0.16	0.13, 0.76	0.0058
MGMT (Methylated)	0.77	0.20	0.38, 1.17	0.0001
Association Test	-0.19	0.046	-0.28, -0.096	<0.0001
log(shape)	0.42	0.059		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 3
Neurocognitive Battery Composite Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	-1.24	0.25	-1.74, -0.75	<0.0001
Time (Years)	-0.97	0.24	-1.46, -0.79	0.0003
Treatment (Arm 1)	0.16	0.16	-0.16, 0.47	0.33
MGMT (Methylated)	0.033	0.17	-0.31, 0.38	0.85
RPA IV (RPA III)	-0.20	0.22	-0.63, 0.23	0.35
RPA V (RPA III)	-1.13	0.30	-1.71, 0.54	0.0002
B. Joint Model				
Longitudinal Sub-Model				
Intercept	-1.10	0.28	-1.65, -0.55	0.0001
Time (Years)	0.12	0.18	-0.23, 0.46	0.51
Treatment arm (Arm 1)	-0.40	0.18	-0.75, -0.056	0.023
MGMT (Methylated)	0.19	0.19	-0.18, 0.56	0.31
RPA IV (RPA III)	-0.55	0.24	-1.03, -0.076	0.023
RPA V (RPA III)	-1.77	0.33	-2.41, -1.12	<0.0001
Survival Sub-Model				
Intercept	-1.22	0.20	-1.62, -0.83	<0.0001
Treatment arm (Arm 1)	-0.14	0.11	-0.35, 0.068	0.19
RPA IV (RPA III)	0.66	0.13	0.42, 0.91	<0.0001
RPA V (RPA III)	0.45	0.16	0.13, 0.76	0.0056
MGMT (Methylated)	0.74	0.21	0.33, 1.14	0.0003
Association Test	-0.11	0.037	-0.19, -0.042	0.0020
log(shape)	0.27	0.042		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects in the mixed model but the random effect for time was removed from the joint model due to convergence issues.

Supplemental Table 4
Cognitive Factor Score – Square Root Transformation

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	0.89	0.098	0.69, 1.08	<0.0001
Time (Weeks)	0.0032	0.0017	-0.00020, 0.0066	0.068
Treatment (Arm 1)	-0.088	0.062	-0.21, 0.034	0.16
Treatment*Time	0.0048	0.0023	0.00040, 0.0093	0.034
MGMT (Methylated)	-0.030	0.065	-0.16, 0.098	0.64
RPA IV (RPA III)	0.069	0.086	-0.10, 0.24	0.43
RPA V (RPA III)	0.47	0.11	0.26, 0.68	<0.0001
B. Joint Model				
Longitudinal Sub-Model				
Intercept	0.93	0.14	0.65, 1.20	<0.0001
Time (Weeks)	0.0032	0.0018	-0.00025, 0.0066	0.069
Treatment arm (Arm 1)	-0.097	0.062	-0.23, 0.025	0.12
Treatment*Time	0.0053	0.0023	0.00078, 0.0098	0.022
MGMT (Methylated)	-0.032	0.064	-0.16, 0.094	0.62
RPA IV (RPA III)	0.071	0.086	-0.097, 0.24	0.41
RPA V (RPA III)	0.47	0.11	0.26, 0.67	<0.0001
Survival Sub-Model				
Intercept	-6.14	0.51	-7.14, -5.14	<0.0001
Treatment arm (Arm 1)	-0.19	0.10	-0.39, 0.015	0.069
RPA IV (RPA III)	0.56	0.12	0.33, 0.79	<0.0001
RPA V (RPA III)	0.48	0.16	0.17, 0.80	0.0025
MGMT (Methylated)	0.84	0.19	0.46, 1.22	<0.0001
Association Test	0.13	0.095	-0.055, 0.32	0.17
log(shape)	0.24	0.041		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 5
Composite Symptom Score – Square Root Transformation

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	1.12	0.068	0.99, 1.25	<0.0001
Time (Years)	1.12	0.21	0.70, 1.54	<0.0001
Treatment (Arm 1)	-0.061	0.042	-0.14, 0.021	0.14
$Time^2$	0.25	0.086	0.081, 0.42	0.0040
$Time^3$	-3.39	0.62	-4.60, -2.17	<0.0001
MGMT (Methylated)	2.35	0.47	1.43, 3.27	<0.0001
RPA IV (RPA III)	-0.04	0.045	-0.13, 0.045	0.34
RPA V (RPA III)	-0.095	0.059	-0.21, 0.022	0.11
B. Joint Model				
Longitudinal Sub-Model				
Intercept	1.15	0.099	0.96, 1.34	<0.0001
Time (Years)	1.23	0.23	0.79, 1.68	<0.0001
Treatment arm (Arm 1)	-3.62	0.65	-4.90, -2.34	<0.0001
$Time^2$	2.50	0.49	1.54, 3.46	<0.0001
$Time^3$	-0.059	0.042	-0.14, 0.023	0.16
MGMT (Methylated)	0.23	0.087	0.062, 0.40	0.0077
RPA IV (RPA III)	-0.042	0.045	-0.13, 0.045	0.35
RPA V (RPA III)	-0.093	0.059	-0.21, 0.022	0.11
Survival Sub-Model				
Intercept	-0.94	0.47	-1.87, -0.0090	0.048
Treatment arm (Arm 1)	-0.16	0.10	-0.36, 0.040	0.12
RPA IV (RPA III)	0.57	0.12	0.34, 0.80	<0.0001
RPA V (RPA III)	0.47	0.16	0.16, 0.79	0.0031
MGMT (Methylated)	0.90	0.19	0.53, 1.27	<0.0001
Association Test	-0.061	0.091	-0.24, 0.12	0.50
log(shape)	0.29	0.064		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 6
Neurologic Factor Score – Square Root Transformation

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	0.80	0.091	0.62, 0.97	<0.0001
Time (Weeks)	-0.0013	0.0018	-0.0048, 0.0023	0.49
Treatment (Arm 1)	-0.14	0.060	-0.25, -0.020	0.022
Treatment*Time	0.0048	0.0024	0.00020, 0.0095	0.042
MGMT (Methylated)	0.0029	0.060	-0.12, 0.12	0.96
RPA IV (RPA III)	-0.066	0.079	-0.22, 0.090	0.41
RPA V (RPA III)	0.20	0.099	0.0083, 0.40	0.041
B. Joint Model				
Longitudinal Sub-Model				
Intercept	0.80	0.14	0.53, 1.08	<0.0001
Time (Weeks)	-0.0011	0.0018	-0.0047, 0.0024	0.53
Treatment arm (Arm 1)	-0.16	0.060	-0.28, -0.043	0.0075
Treatment*Time	-0.0077	0.060	0.00084, 0.010	0.90
MGMT (Methylated)	-0.051	0.86	-0.12, 0.11	0.56
RPA IV (RPA III)	0.22	0.11	-0.22, 0.12	0.037
RPA V (RPA III)	0.0055	0.0024	0.014, 0.43	0.021
Survival Sub-Model				
Intercept	-6.26	0.52	-7.28, -5.24	<0.0001
Treatment arm (Arm 1)	-0.22	0.11	-0.42, -0.0066	0.043
RPA IV (RPA III)	0.58	0.12	0.34, 0.81	<0.0001
RPA V (RPA III)	0.52	0.16	0.20, 0.84	0.0013
MGMT (Methylated)	0.80	0.19	0.42, 1.18	<0.0001
Association Test	0.32	0.12	0.089, 0.55	0.0067
log(shape)	0.26	0.041		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 7
Treatment Factor Score – Square Root Transformation

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	2.01	0.23	1.56, 2.46	<0.0001
Time (Years)	7.77	0.93	5.95, 9.59	<0.0001
Treatment (Arm 1)	-0.077	0.14	-0.35, 0.20	0.58
$Time^2$	1.00	0.34	0.33, 1.67	0.0039
$Time^3$	-21.95	2.68	-27.21, -16.69	<.0001
MGMT (Methylated)	14.73	2.03	10.75, 18.71	<.0001
RPA IV (RPA III)	-0.41	0.15	-0.70, -0.12	0.0057
RPA V (RPA III)	-0.31	0.20	-0.70, 0.073	0.11
B. Joint Model				
Longitudinal Sub-Model				
Intercept	1.26	0.13	1.019, 1.51	<0.0001
Time (Years)	3.51	0.34	2.84, 4.18	<0.0001
Treatment arm (Arm 1)	-9.64	0.99	-11.59, -7.70	<0.0001
$Time^2$	6.40	0.76	4.91, 7.88	<0.0001
$Time^3$	-0.0097	0.054	-0.12, 0.096	0.86
MGMT (Methylated)	0.30	0.13	0.048, 0.56	0.020
RPA IV (RPA III)	-0.14	0.056	-0.25, -0.029	0.014
RPA V (RPA III)	-0.10	0.074	-0.24, 0.050	0.18
Survival Sub-Model				
Intercept	-0.60	0.48	-1.54, 0.35	0.21
Treatment arm (Arm 1)	-0.17	0.10	-0.37, 0.030	0.095
RPA IV (RPA III)	0.55	0.12	0.32, 0.78	<0.0001
RPA V (RPA III)	0.46	0.16	0.15, 0.77	0.0040
MGMT (Methylated)	0.89	0.19	0.52, 1.27	<0.0001
Association Test	-0.11	0.046	-0.20, -0.016	0.021
log(shape)	0.41	0.072		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 8
Composite Interference Score – Square Root Transformation

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	1.21	0.10	1.02, 1.41	<0.0001
Time (Years)	1.05	0.35	0.37, 1.73	0.0024
Treatment (Arm 1)	-0.021	0.064	-0.15, 0.11	0.74
$Time^2$	0.43	0.14	0.17, 0.70	0.0017
$Time^3$	-4.35	1.00	-6.30, -2.39	<0.0001
MGMT (Methylated)	3.21	0.75	1.73, 4.69	<0.0001
RPA IV (RPA III)	-0.0082	0.066	-0.14, 0.12	0.90
RPA V (RPA III)	-0.15	0.087	-0.32, 0.021	0.086
B. Joint Model				
Longitudinal Sub-Model				
Intercept	1.28	0.14	1.00, 1.57	<0.0001
Time (Years)	-0.031	0.26	-0.53, 0.47	0.90
Treatment arm (Arm 1)	-1.37	0.75	-2.84, 0.10	0.068
$Time^2$	1.27	0.61	0.069, 2.48	0.038
$Time^3$	-0.046	0.064	-0.17, 0.079	0.47
MGMT (Methylated)	0.50	0.14	0.23, 0.48	0.0003
RPA IV (RPA III)	-0.0093	0.065	-0.14, 0.12	0.89
RPA V (RPA III)	-0.14	0.085	-0.31, 0.024	0.094
Survival Sub-Model				
Intercept	-1.09	0.49	-2.04, -0.13	0.025
Treatment arm (Arm 1)	-0.42	0.13	-0.66, -0.17	0.0008
RPA IV (RPA III)	0.63	0.13	0.38, 0.87	<0.0001
RPA V (RPA III)	0.62	0.17	0.29, 0.96	0.0003
MGMT (Methylated)	0.79	0.20	0.39, 1.19	0.0001
Association Test	0.56	0.11	0.34, 0.78	<0.0001
log(shape)	0.29	0.068		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 9
Global Quality of Life (QOL) Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	70.78	2.66	65.56, 76.00	<0.0001
Time (Years)	-41.23	9.23	-59.35, -23.12	<0.0001
Treatment (Arm 1)	-1.21	1.59	-4.34, 1.92	0.45
$Time^2$	133.55	26.91	80.75, 186.34	<0.0001
$Time^3$	-95.13	20.31	-134.97, -55.28	<0.0001
MGMT (Methylated)	0.047	1.73	-3.36, 3.45	0.98
RPA IV (RPA III)	0.22	2.29	-4.27, 4.71	0.92
RPA V (RPA III)	-8.04	2.88	-13.70, -2.39	0.0054
B. Joint Model				
Longitudinal Sub-Model				
Intercept	71.29	2.62	66.16, 76.42	<0.0001
Time (Years)	-52.14	8.84	-69.46, -34.82	<0.0001
Treatment arm (Arm 1)	164.80	25.90	114.05, 215.56	<0.0001
$Time^2$	-116.42	19.79	-155.19, -77.64	<0.0001
$Time^3$	-1.26	1.56	-4.32, 1.79	0.42
MGMT (Methylated)	0.071	1.69	-3.24, 3.38	0.97
RPA IV (RPA III)	0.26	2.24	-4.13, 4.65	0.91
RPA V (RPA III)	-7.88	2.82	-13.42, -2.34	0.0053
Survival Sub-Model				
Intercept	-1.17	0.45	-2.05, -0.29	0.0090
Treatment arm (Arm 1)	-0.18	0.10	-0.39, 0.016	0.071
RPA IV (RPA III)	0.56	0.12	0.33, 0.79	<0.0001
RPA V (RPA III)	0.48	0.16	0.16, 0.79	0.0028
MGMT (Methylated)	0.92	0.19	0.54, 1.29	<0.0001
Association Test	0.0065	0.0029	0.00087, 0.012	0.024
log(shape)	0.39	0.073		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 10
Cognitive Function Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	75.37	3.01	69.45, 81.28	<0.0001
Time (Weeks)	0.034	0.058	-0.080, 0.15	0.56
Treatment (Arm 1)	1.42	1.94	-2.38, 5.23	0.46
Treatment*Time	-0.19	0.075	-0.34, -0.044	0.011
MGMT (Methylated)	0.68	1.98	-3.21, 4.58	0.73
RPA IV (RPA III)	-0.054	2.63	-5.22, 5.11	0.98
RPA V (RPA III)	-13.48	3.27	-19.92, -7.05	<0.0001
B. Joint Model				
Longitudinal Sub-Model				
Intercept	73.69	4.64	64.59, 82.79	<0.0001
Time (Weeks)	0.035	0.058	-0.079, 0.15	0.55
Treatment arm (Arm 1)	1.83	2.00	-2.09, 5.74	0.36
Treatment*Time	-0.23	0.076	-0.37, -0.078	0.0027
MGMT (Methylated)	1.10	2.09	-2.99, 5.19	0.60
RPA IV (RPA III)	-0.12	2.83	-5.65, 5.42	0.97
RPA V (RPA III)	-13.21	3.43	-19.94, -6.48	0.0001
Survival Sub-Model				
Intercept	-5.54	0.60	-6.71, -4.36	<0.0001
Treatment arm (Arm 1)	-0.20	0.11	-0.42, 0.0080	0.059
RPA IV (RPA III)	0.60	0.12	0.37, 0.84	<0.0001
RPA V (RPA III)	0.51	0.16	0.20, 0.83	0.0015
MGMT (Methylated)	0.85	0.19	0.47, 1.23	<0.0001
Association Test	-0.0086	0.0034	-0.015, -0.0019	0.011
log(shape)	0.27	0.041	-	<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 11
Motor Dysfunction Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	10.91	2.34	6.32, 15.51	<0.0001
Time (Weeks)	0.015	0.042	-0.067, 0.098	0.71
Treatment (Arm 1)	-2.71	1.47	-5.60, 0.19	0.067
Treatment*Time	0.11	0.055	-0.0011, 0.21	0.053
MGMT (Methylated)	-1.03	1.55	-4.08, 2.01	0.51
RPA IV (RPA III)	3.68	2.06	-0.37, 7.72	0.075
RPA V (RPA III)	20.02	2.56	14.99, 25.04	<0.0001
B. Joint Model				
Longitudinal Sub-Model				
Intercept	11.90	3.42	5.19, 18.60	0.0005
Time (Weeks)	-3.23	1.47	-6.10, -0.35	0.028
Treatment arm (Arm 1)	0.015	0.043	-0.069, 0.099	0.73
Treatment*Time	0.15	0.055	0.042, 0.26	0.0064
MGMT (Methylated)	-0.82	1.53	-3.82, 2.17	0.59
RPA IV (RPA III)	3.63	2.10	-0.48, 7.74	0.083
RPA V (RPA III)	20.15	2.63	14.98, 25.31	<0.0001
Survival Sub-Model				
Intercept	-6.24	0.53	-7.27, -5.21	<0.0001
Treatment arm (Arm 1)	-0.28	0.11	-0.49, -0.059	0.013
RPA IV (RPA III)	0.61	0.12	0.37, 0.85	<0.0001
RPA V (RPA III)	0.44	0.16	0.11, 0.76	0.0080
MGMT (Methylated)	0.49	0.21	0.071, 0.91	0.022
Association Test	0.022	0.0042	0.04, 0.030	<0.0001
log(shape)	0.28	0.043		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.

Supplemental Table 12
Communication Deficit Score

	Estimate	Std. Error	95% CI	p-value
A. Mixed Effects Model				
Intercept	13.07	2.83	7.51, 18.63	<0.0001
Time (Years)	-1.59	2.13	-5.78, 2.61	0.46
Treatment (Arm 1)	-3.95	1.76	-7.42, -0.49	0.026
Treatment*Time	8.42	2.79	2.91, 13.93	0.0030
MGMT (Methylated)	1.34	1.88	-2.36, 5.04	0.48
RPA IV (RPA III)	2.53	2.50	-2.38, 7.45	0.31
RPA V (RPA III)	13.15	3.10	7.05, 19.24	<0.0001
B. Joint Model				
Longitudinal Sub-Model				
Intercept	13.69	2.62	8.55, 18.82	<0.0001
Time (Years)	-1.60	2.12	-5.76, 2.56	0.45
Treatment arm (Arm 1)	-3.92	1.49	-6.84, -0.99	0.0086
Treatment*Time	9.22	2.80	3.73, 14.71	0.0010
MGMT (Methylated)	1.04	1.71	-2.32, 4.40	0.54
RPA IV (RPA III)	1.50	0.59	0.33, 2.66	0.012
RPA V (RPA III)	11.70	2.30	7.18, 16.21	<0.0001
Survival Sub-Model				
Intercept	-1.02	0.44	-1.89, -0.16	0.020
Treatment arm (Arm 1)	-0.18	0.10	-0.38, 0.019	0.076
RPA IV (RPA III)	0.56	0.12	0.33, 0.79	<0.0001
RPA V (RPA III)	0.47	0.16	0.15, 0.78	0.0034
MGMT (Methylated)	0.82	0.20	0.43, 1.20	<0.0001
Association Test	0.0046	0.0031	-0.0015, 0.011	0.14
log(shape)	0.25	0.040		<0.0001

MGMT=Hypermethylation of the O^6-methylguanine-DNA-methyltransferase; RPA=Recursive Partitioning Analysis; CI=Confidence interval; Reference Category/Level is specified in brackets (). Intercept and time are random effects.