Fine Motor Skill Decline after Brain RT – A Multivariate Normal Tissue Complication Probability Study of a Prospective Trial

Multivariate NTCP Analysis of Motor Decline

46 Characters or Fewer Including Spaces

Michael Connor, MD\textsuperscript{a}
Mia Salans, MD, MAS\textsuperscript{a}
Roshan Karunamuni, PhD\textsuperscript{a}
Soumya Unnikrishnan, MD\textsuperscript{a}
Minh-Phuong Huynh-Le, MD, MAS\textsuperscript{b}
Michelle Tibbs, MD, MAS\textsuperscript{a}
Alexander Qian, MD\textsuperscript{a}
Anny Reyes, PhD\textsuperscript{c}
Alena Stasenko, PhD\textsuperscript{c}
Carrie McDonald, PhD\textsuperscript{a,c}
Vitali Moiseenko, PhD\textsuperscript{a}
Issam El-Naqa, PhD\textsuperscript{d}
Jona Hattangadi-Gluth, MD\textsuperscript{a}

\textsuperscript{a}Department of Radiation Medicine and Applied Sciences, University of California San Diego, San Diego, CA, USA
\textsuperscript{b} Radiation Oncology, The Brooklyn Cancer Center, Brooklyn, NY, USA
\textsuperscript{c}Department of Psychiatry, University of California San Diego, San Diego, CA, USA
\textsuperscript{d}Department of Radiation Oncology, Moffitt Cancer Center and Research Institute, Tampa, FL, USA

Corresponding Author:
Jona Hattangadi-Gluth

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Department of Radiation Medicine and Applied Sciences
University of California, San Diego
9500 Gilman Dr. Mail Code 0861
La Jolla, CA 92093-0861
Phone: 858-822-6040
Email: jhattangadi@health.ucsd.edu
Twitter: @JHGLab

Author Responsible for Statistical Analysis Name & Email Address:
Michael Connor
Department of Radiation Medicine and Applied Sciences
University of California, San Diego
9500 Gilman Dr. Mail Code 0861
La Jolla, CA 92093-0861
m3connor@health.ucsd.edu

Conflict of Interest Statement for All Authors:
CRM reports research funding from GE Healthcare, unrelated to the current study. There are no other financial or other relationships that might lead to a perceived conflict of interest.

Meeting Presentations: Analyses from this study were previously presented as an oral presentation at the 2021 annual American Society of Radiation Oncology (ASTRO) meeting

Funding Statement:
This work was supported by the Radiological Society of North America (RSNA) (Research Resident Grant to MC), the National Institutes of Health (1TL1TR001443 to MS, MT; F31 NS111883-01 to AR; UL1TR001442 of CTSA funding in support of CTRI; and 1KL2TR001444, UL1TR000100, R01 CA238783-01 to JAH-G); National Cancer Institute and UC San Diego Moores Cancer Center (P30 CA02310029 to JAH-G); and American Cancer Society (RSG-15-229-01-CCE
to CRM). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies. Funding agencies had no direct role in designing, conducting, or reporting the study.

Data Availability Statement for this Work
Research data are stored in an institutional repository and will be shared upon request to the corresponding author

Acknowledgements
Special thanks to the patients who enrolled and participated in this study. Many thanks to nurses in radiation oncology, particularly Abby Pennington, RN and Mary Kay Gorman, RN, for their help with this study.
Abstract

Purpose
Brain radiotherapy can impair fine motor skills (FMS). FMS are essential for activities of daily living, enabling hand-eye coordination for manipulative movements. We developed normal tissue complication probability (NTCP) models for decline in FMS after fractionated brain RT.

Methods
On a prospective trial, 44 primary brain tumor patients received fractioned RT, underwent high-resolution volumetric MRI and diffusion tensor imaging, and comprehensive FMS assessments (Delis-Kaplan Executive Function System Trail Making Test Motor Speed [DKEFS-MS], and Grooved Pegboard Dominant/Non-Dominant Hands [PDH/PNDH]) at baseline and 6 months post-RT. Regions of interest subserving motor function (including cortex, superficial white matter, thalamus, basal ganglia, and white matter tracts) were autosegmented using validated methods and manually verified. Dosimetric and clinical variables were included in multivariate NTCP models, using automated bootstrapped logistic regression, least absolute shrinkage and selection operator (LASSO) logistic regression, and random forests with nested cross-validation.

Results
Half of patients showed decline on PNDH, 17 of 42 (40.4%) on PDH, and 11 of 44 (25%) on DKEFS-TM. Automated bootstrapped logistic regression selected a one-term model including maximum dose to dominant postcentral white matter. LASSO selected this term and steroid use. The top five variables in random forest were all dosimetric: mean and maximum dose to dominant corticospinal tract; maximum dose to dominant thalamus; mean dose to dominant caudate; maximum dose to dominant postcentral white matter. This technique performed best with AUC 0.69 (95% CI 0.68 – 0.70) on nested cross-validation.
Conclusion

We present the first NTCP models for FMS impairment after brain RT. Dose to several supratentorial motor-associated ROIs correlated with decline in dominant hand fine motor dexterity in primary brain tumor patients, outperforming clinical variables. These data can guide prospective fine motor-sparing strategies for brain RT.
Introduction

Intracranial radiotherapy (RT) is associated with neurocognitive decline, likely due to injury to eloquent neuroanatomic structures\(^1\). As survival for brain tumor patients improves, the burden of treatment sequelae increases, with many of these patients exhibiting dysfunction that is progressive and disabling\(^2\). Although affected neurocognitive domains such as learning, memory, processing speed, attention, and executive function often garner the most interest\(^3\), fine motor skills (FMS) can also be impaired, with randomized trials demonstrating decline in this domain after RT\(^4\)–\(^6\).

FMS involve precise hand-eye coordination for manipulative movements and are essential for many activities of daily living, including handwriting, typing, eating, and using a phone. In addition, FMS are important for performance on other tests of cognitive function\(^7\). With increased focus on survivorship and quality of life, strategies to improve or preserve these outcomes may consider FMS as well as higher-order cognitive domains. Novel imaging and image processing techniques allow us to well demarcate eloquent brain regions quickly and reliably over time and understand how radiation dose-related damage may underlie functional impairments\(^4,8\)–\(^11\).

While neurocognitive decline is a common sequalae of RT for brain tumor patients, there is little in the way of predictive models to guide preservation strategies in radiotherapy patients. We previously reported seminal work on changes in FMS after brain RT, examining \textit{in vivo} imaging biomarkers of white matter integrity and cortical atrophy in motor-associated regions of interest as predictors of longitudinal FMS decline\(^4\). This study established associations between radiation dose and loss of cortical and white matter integrity, as well as associations between imaging biomarkers of injury and functional FMS impairment. While we...
successfully demonstrated the associations between RT dose and microstructural damage within the motor cortex and superficial white matter, normal tissue complication probability (NTCP) analyses would be needed for a more practical, clinical model directly exploring the associations between RT dose and functional outcome. These are the very dose/volume/structure parameters needed to design fine motor functional-sparing interventions. Very few prior reports, mostly pertaining to the hippocampus\textsuperscript{12}, have described an NTCP dose response model for a cognitive or functional outcome after brain RT, and QUANTEC (Quantitative Analysis of Normal Tissue Effects in the Clinic) guidelines do not include cognitive or motor outcomes. We sought to perform multivariate NTCP analyses to identify predictors of FMS decline at 6-months post-RT in primary brain tumor patients undergoing fractionated partial brain RT, using data from a prospective clinical trial. Our methodology uses advanced image processing and a cognitive/fine motor neuroscience framework. Given the multiple regions which subserve motor function, and the multiple clinical variables which can affect function, we used multivariate model building techniques. We specifically analyzed baseline to 6-month change in FMS as this approximates the shift from subacute to long-term, irreversible RT-associated damage\textsuperscript{2,8}.

Methods

Study Overview

We enrolled 52 patients on a single-institution prospective longitudinal IRB-approved clinical trial investigating comprehensive neurocognitive functioning in several domains, including FMS, in patients receiving fractionated RT for primary brain tumors. Patients underwent high resolution volumetric MRI and diffusion imaging, as well as comprehensive neurocognitive evaluation, at
baseline (pre-RT) and 3-, 6-, and 12 months after RT completion. All subjects provided written informed consent. Inclusion criteria included: Karnofsky performance status [KPS] ≥70; diagnosis of primary brain tumor; life expectancy of ≥1 year; ability to complete neurocognitive assessments in English; and age ≥18 years. Patients who received prior radiation were excluded. The current analysis includes 44 of these primary brain tumor patients with baseline and 6-month FMS outcomes as well as baseline imaging with the specific MRI protocol detailed below.

**FMS Assessment**

FMS was evaluated using two robust, validated objective measurements. All FMS assessments were performed in-person, with direct observation and testing by a trained neuropsychologist. Fine motor speed was evaluated with the Delis-Kaplan Executive Functioning System Trail Making Test Motor Speed\(^{13}\) (DKEFS-MS) and fine motor coordination/dexterity was evaluated with the Grooved Pegboard\(^{14}\) test. In the DKEFS-MS test, participants are instructed to connect a series of circles joined by a dashed line as quickly as possible. Scores are defined by the time taken to complete the test in seconds. The Grooved Pegboard requires patients to insert metal pegs into slots in sequence as quickly as possible. The score is the time in seconds required to complete the array with the dominant (PDH) and non-dominant (PNDH) hands, with a higher score indicating worse performance.

**Reliable Change Indices**

Reliable change indices (RCI), quantifications of whether the change in FMS scores per test is
significant for individual patients, were calculated between baseline and 6-month scores using the standardized neurocognitive scores (T-scores)\textsuperscript{15}. To account for repeated testing, reliable change indices were adjusted for practice effects (RCI-PEs) using DKEFS-MS T-scores, corrected for age, sex, and education when appropriate, and raw PDH and PNDH scores. RCI-PEs were calculated for each patient on each neurocognitive test measure between 0 and 6 months. RCI-PE calculation is based on the test-retest data of a reference group that has completed the same assessment multiple times\textsuperscript{16,17}. A negative RCI-PE was scored as decline.

**Imaging**

The imaging acquisition of high-resolution volumetric and diffusion-weighted MRIs for this study have been described in detail elsewhere\textsuperscript{18,19}. Briefly, imaging for all patients at each time point were acquired on a 3.0T 750 GE system (GE Healthcare, Milwaukee, Wisconsin) equipped with an 8-channel head coil. Sequences selected for the protocol included a 3D volumetric T1-weighted inversion recovery spoiled gradient echo sequence (echo time [TE]/repetition time [TR]=2.8/6.5 ms; inversion time [TI]=450 ms; flip angle=8 degrees; field of view [FOV]=24 cm), a 3D FLAIR sequence (TE/TR=125/6000 ms, TI=1868 ms, FOV=24 cm, matrix=256x256, slice thickness=1 mm), and a diffusion weighted imaging (DWI) sequence using a single-shot pulsed-field gradient spin EPI sequence (TE/TR=96 ms/17 s; FOV=24 cm, matrix=128x128x48; in-plane resolution 1.87x1.875; slice thickness=2.5 mm; 48 slices) with b=0, 500, 1500, and 4000 s/mm\textsuperscript{2}, with 1, 6, 6, and 15 unique gradient directions for each b-value respectively and one average for each non-zero b-value. Two additional b=0 volumes were acquired with either forward or reverse phase-encode polarity for use in nonlinear B\textsubscript{0} distortion correction\textsuperscript{20}.

**Image Analysis and Segmentation**
Imaging data was preprocessed using in-house algorithms in MATLAB. T1- and T2-weighted images were corrected for distortions attributed to gradient nonlinearities\textsuperscript{21} and imaging non-uniformities arising from bias fields\textsuperscript{22}. Diffusion-weighted data was corrected for spatial distortions arising from eddy currents, and B0 field inhomogeneities using robust, well-validated methods\textsuperscript{20}.

We specifically investigated brain regions which subserve FMS. As described previously\textsuperscript{4}, selected FMS regions of interest (ROIs) included the sensorimotor cortex and superficial white matter (WM) (precentral, postcentral, and paracentral gyri), the corticospinal tracts, the cerebellar cortex and WM, the thalamus, and the basal ganglia (caudate, putamen, and pallidum) (Table S1). ROIs were classified as dominant or non-dominant side per subject based on handedness. For example, in a right-handed patient, the left corticospinal tract was the dominant one. Cortical, superficial WM, cerebellar, thalamic, and basal ganglia ROIs were segmented using the automated FreeSurfer processing pipeline (version 5.3; http://surfer.nmr.harvard.edu), available on the Neuroscience Gateway Portal\textsuperscript{23}. Superficial WM is defined as the volume of WM up to 5 mm below the surface of cortical ROIs defined by FreeSurfer's Desikan-Killiany atlas\textsuperscript{24}. The DWI-derived maps and high-resolution volumetric MRI were co-registered and atlas-based tractography was used to segment the DWI into WM tracts in order to evaluate the corticospinal tracts\textsuperscript{25}. A censoring mask was drawn manually, slice by slice, on each image to exclude tissue affected by tumor, surgical cavity, or edema. Voxels within the censoring mask were excluded from the final ROI to avoid confounding by tumor and edema-related effects\textsuperscript{26}. Planning CT and RT dose maps were co-registered to the baseline T1 and DWI volumes to estimate dose distributions for each of the ROIs\textsuperscript{19}.

**Candidate Variables**
The mean and maximum doses to all of the above ROIs (Table S1) were calculated and included as potential variables. Volumetric dose variables (e.g., V10Gy) were included in univariate analyses, however they were excluded from multivariate model building due to the small volume of most ROIs and high collinearity\textsuperscript{27,28}. Additionally, several clinical and demographic variables were included, all binary or continuous: age (years), prescription dose (Gy), receipt of adjuvant chemotherapy (yes/no), anti-epileptic drug use (yes/no), diagnosis (glioma vs. other), ethnicity (Hispanic vs. non-Hispanic), handedness (L vs. R), Karnofsky performance status, KPS (≥90 or ≤80), laterality (tumor on dominant or non-dominant side), progression at 6 months (yes/no), race (white vs. other), radiation modality (photon vs. proton), seizures (yes/no), sex (M/F), steroid use (yes/no), any surgery, and GTR (gross total resection) at surgery.

**Statistical Analyses**

Statistical analysis was done using R\textsuperscript{29} with packages including mlr\textsuperscript{30}, caret\textsuperscript{31}, and glmnet\textsuperscript{32}.

**Univariate Analyses**

Univariate analyses were performed using logistic regression, Spearman’s rank correlation (Rs), non-parametric tests (Wilcoxon rank test for continuous variables, Fisher’s test for categorical variables), and decision stumps. A decision stump is a one-level decision tree, with a single root node connected immediately to the terminal nodes, and makes a decision based on a single input feature.

**Multivariate Analyses**

Multivariate model building was performed using three commonly used methods in this space\textsuperscript{33}: automated bootstrapped logistic regression with forward selection, LASSO, and random forests. First, in order to minimize issues with collinearity, among variables highly correlated with each
other (Pearson’s correlation coefficient larger than 0.85), those with the lower correlation with
the outcome were removed.

The automated logistic regression technique has been described previously, first introduced by
El Naqa and colleagues\textsuperscript{34}. Briefly, the modeling process is done in two steps. First, an optimal
model size (or number of variables included) is estimated by automating forward selection over
1000 bootstrapped samples. The average predictive performance on each out of bag sample for
each model order is calculated, and the peak performance of this statistic (e.g., Spearman rank
correlation) as a function of the number of variables in the model is used to select the optimal
model size. Second, models of this optimal order are fit across the held-out folds of a repeated
cross-validation (5-folds, 200 repeats), and the most-frequently selected variable set is chosen
for the final model.

LASSO, or least absolute shrinkage and selection operator, is a regression technique with a
penalty parameter. Specifically, L1 regularization adds a penalty equal to the absolute value of
the coefficients, where \textit{lambda} is a tuning parameter controlling the amount of regularization.
LASSO therefore performs feature selection, encourages sparsity of model parameters, and is
well-suited for multicollinearity.

Random Forests are ensembles of many decision trees; the algorithm utilizes both bagging and
random subsets of features to create an uncorrelated “forest” of these trees, reducing
overfitting. The tunable hyperparameter was the number of variables considered as candidate
splitting variables at each split when building each tree, or \textit{mtry} in the randomForest R
package\textsuperscript{35}. Variables were ranked by mean decrease in the Gini index; the higher the value the
higher the importance of the variable in the model\textsuperscript{33}. 
Nested cross validation with a 5-fold inner loop for tuning hyperparameters was done; the outer loop also consisting of 5-folds with the held-out data used as a test set for unbiased performance estimation. In each inner loop for the LASSO, an initial feature selection of the top 4 variables by AUC using each feature separately for thresholded class prediction was also implemented to further reduce the feature space and tendency for overfitting, due to poor performance noted without this. Outer cross-validation with random stratified reshuffling was repeated 200 times. Final models were fit by applying optimal hyperparameters and training on the full dataset.

Model performance, discrimination, and calibration were assessed by the area under the ROC curve (AUC), balanced accuracy, Brier's score, Nagelke's R2, calibration slopes and intercepts, and Hosmer-Lemeshow tests.

Results

Patient Characteristics and FMS Outcomes

Patient characteristics are shown in Table 1. In total, 44 primary brain tumor patients were eligible for inclusion in this study with baseline and 6-month FMS outcomes and baseline volumetric and diffusion imaging. Of these, all 44 completed baseline and six-month DKEFS-TM tests, and 42 completed pegboard tests. The median age was 46.2 years, 57% were male, and 89% were right-handed. The cohort was high-functioning with 93% having a KPS ≥ 90. Most patients (61%) had gliomas, and 13 (29.5%) had benign diagnoses. By RCI-PE, 21 of 42 patients (50%) experienced decline on the PNDH test at 6 months, 11 of 44 (25%) experienced decline on the DKEFS-TM test, and 17 of 42 (40.4%) declined on the PDH.
Univariate Analyses

On univariate analysis for PDH outcome at 6 months, mean or maximum dose to most dominant supratentorial structures, as well as several of their volumetric dose variables (e.g., V30Gy, V40Gy) were correlated with decline (Table 2, Fig 1). Non-dominant structures showed no association. Using decision stumps and in-sample performance, $D_{\text{max}}$ to the precentral cortex, $D_{\text{mean}}$ to the caudate, and $D_{\text{mean}}$ to the thalamus were the most discriminative dosimetric variables (AUC 0.76-0.77), with cutoff doses of 24.3 Gy, 33.7 Gy, and 27.3 Gy, respectively.

Among the clinical variables, laterality (presence of tumor on the dominant side) and steroid use were correlated with decline. There were no significant associations for decline on the non-dominant pegboard test (Table S1), and only increasing age was predictive for decline on the DKEFS-TM (Table S2). Dosimetric variables among dominant motor ROIs showed a high degree of correlation (Fig S1).

Multivariate Analyses

We proceeded with multivariate model building for decline on the PDH at 6 months. Automated bootstrapped logistic regression selected an optimal model order of 1, likely reflecting overfitting with higher order models due to our small sample size. The most frequently selected variable was maximum dose to the dominant postcentral white matter, with the next two most frequent being steroid use, and mean dose to the dominant caudate (Fig 2A). Model performance at nested cross-validation by the AUC as 0.61 (SD 0.005) (Fig 3). The final NTCP model is given by:

$$NTCP = \frac{1}{1 + e^{-S}}, \text{where } S = -1.47 + (0.04 \times \text{dom\_wm\_postcentral\_}D_{\text{max}})$$

LASSO model performance at nested cross validation by AUC was 0.55 (SD 0.005), improved to 0.63 (SD 0.005) by the use of an initial feature selection step based on AUC (Fig 3). The
frequency of terms selected over the repeated, nested CVs are shown in Fig 2A. The final
LASSO NTCP model (without feature selection) is given by:

\[ NTCP = \left(1 + e^{-S}\right)^{-1}, \text{where } S = -0.81 + (0.013 \times \text{dom}_\text{wm}_\text{postcentral}_\text{Dmax}) + (0.31 \times \text{steroid}_\text{use}[0 = \text{yes}; 1 = \text{no}]) \]

The final LASSO NTCP model (with feature selection) is given by:

\[ NTCP = \left(1 + e^{-S}\right)^{-1}, \text{where } S = -1.44 + (0.025 \times \text{dom}_\text{wm}_\text{postcentral}_\text{Dmax}) + (0.94 \times \text{steroid}_\text{use}[0 = \text{no}; 1 = \text{yes}]) \]

Graphical depictions of the logistic NTCP curves are shown in the supplement (Fig S3, S4).

The top five important variables in the Random Forest, as ranked by mean decrease in Gini
coefficient, were: mean dose to dominant corticospinal tract; maximum dose to the dominant
thalamus; mean dose to dominant corticospinal tract; mean dose to the dominant caudate, and
maximum dose to the dominant postcentral white matter (Fig 2B). All of these dosimetric
variables outperformed clinical variables. Model performance by AUC was 0.69 (95% CI 0.68 –
0.70) on nested cross-validation (Fig 3).

All models achieved good calibration by Hosmer-Lemeshow test (p=0.07-0.67). Additional
model characteristics are shown in Table 3.

Multivariate analysis was attempted for the other endpoints, though models performed poorly
via nested cross validation for predicting decline on the PNDH (AUC 0.31-0.49) and DKEFS-TM
(AUC 0.51-0.57) tests.
Discussion

There is a critical need to understand substructure-dose-function relationships in the brain, and to build practical and clinical models for cognitive and functional outcomes, thereby paving the way for more individualized precision brain RT. We present the first evidence-based NTCP modeling analyses of fine motor function decline after brain RT, and the only study, to our knowledge, of a neurocognitive/functional endpoint outside of hippocampal dose and cognition. There are published NTCP studies for the prediction of memory decline based on hippocampal dose\textsuperscript{37–39}, however one of these failed an external validation attempt\textsuperscript{40} and none use multivariate methods or take into account clinical variables. Given that the probability of a functional change in outcome is typically influenced by both dosimetric and clinical factors, we employed multivariate model building methods\textsuperscript{34}. We also employed a neuroanatomic and motor function framework and advanced image processing to analyze specific brain regions of interest which are associated with fine motor function. We found that dose to several supratentorial motor-associated ROIs correlated with decline on grooved pegboard dominant hand tests at 6 months post-RT in primary brain tumor patients undergoing partial fractionated brain RT. These findings implicate the anatomic-functional pathway of fine motor control: including cortex, superficial white matter, thalamus and basal ganglia, and the corticospinal tract. The results also reinforce neuroanatomic correlation; that dose to brain ROIs on the motor dominant side of the cerebral hemisphere matter most.

Although there is more focus on higher-order cognitive domains in the literature, impairment in fine motor skills has been consistently reported in clinical trials\textsuperscript{4–6}. Chang et al found that 33-50% of patients undergoing SRS vs. SRS plus whole brain RT experienced significant decline on grooved pegboard tests of fine motor skills, the highest rates of decline outside of the Hopkins Verbal Learning Test (HVLT) in the study\textsuperscript{6}. Brown et al, in reporting the results
of Alliance N0574, found similar rates of decline on grooved pegboard tests in patients with 1-3 metastases randomized to SRS with or without whole brain RT, with 29.3-47.7% of patients declining on grooved pegboard tests\textsuperscript{5}. Moreover, fine motor skills are critical for performance on other common neurocognitive instruments used in clinical trials, including tests of processing speed and those that require a writing/drawing component. Some investigators have even used fine motor skill tests as a “measure of higher-level cognitive processes which influence motor and nonmotor skills alike”\textsuperscript{41–43}.

We previously found that increasing age and use of anti-epileptic medications predicted higher rates of fine motor skill decline among primary brain tumor patients undergoing fractionated RT\textsuperscript{4}. We also found imaging biomarkers for FMS decline, showing that diffusion imaging measures of white matter microstructure injury in the cerebellar white matter and corticospinal tract were predictive of decline, as was volumetric atrophy of the precentral and cerebellar cortices, pallidum, and pons\textsuperscript{4}. Likewise, in traumatic brain injury, atrophy of the thalamus, putamen, and pallidum have been correlated significantly with decline in FMS\textsuperscript{44}. Here, we report the results of NTCP modeling to directly predict fine motor skill decline from planned/delivered doses and clinical variables.

Univariate methods showed that mean or max dose to all supratentorial motor ROIs in our study were associated with FMS decline, with a high degree of correlation among variables. Multivariate approaches yielded very parsimonious models likely due to our modest sample size. Multivariate logistic regression techniques (automated bootstrapped logistic regression and LASSO) selected maximum dose to the dominant postcentral white matter and steroid use as the top two features for inclusion in the models. The top five most important predictors in the Random Forest (Fig 2B) were all dosimetric variables relating to dominant-sided supratentorial motor ROIs, all of which outperformed clinical variables. These results reinforce the importance
of dose to motor-associated functional brain areas, with careful attention to lateralization of brain function affecting the dominant hand.

One may expect that dose to the precentral, or primary motor, cortex would be most important in a model predicting fine motor skill decline, and indeed these variables were ranked as highly important in the random forest. Dose variables for the precentral cortex were also significant on univariate analysis (Table 2). Additionally, $D_{\text{max}}$ to the precentral cortex was highly predictive as a decision stump. However, our logistic models selected maximum dose to the postcentral white matter and random forest showed high importance of this region. As an illustrative example, the logistic model predicts increased risk of decline on PDH of 5%, 10%, and 20% over baseline with doses of 6.9 Gy, 13.1 Gy, and 25 Gy, respectively, to the postcentral white matter (Fig. S3). This could represent a nuance of our dataset, as dose to this postcentral area is highly collinear with doses to the precentral cortex given they are neuroanatomically adjacent. However, some studies have demonstrated the importance of injury to the somatosensory cortex to motor function and recovery\textsuperscript{45,46}. Mean dose to caudate on the dominant side, and max dose to thalamus were consistently important dosimetric predictors in our study. These findings are consistent with previous studies of the basal ganglia and thalamus as critical structures in radiosurgery applications. These two areas ranked behind only the pons/midbrain as high-risk regions in a clinical study to predict permanent symptomatic post-SRS injury in AVM patients\textsuperscript{47}. Finally, mean and maximum dose to the dominant corticospinal tract was an important predictor of FMS decline in the dominant hand in our study. This key white matter tract represents a possible (organ at risk) OAR for sparing.

Indeed, there is limited precedent in the literature for implemented motor-associated OARs into treatment planning. Maruyama et al found that a maximum dose of 23 Gy to the corticospinal tract resulted in a 5% complication rate\textsuperscript{48}. Koga et al found that limiting the corticospinal tract to
20 Gy during single fraction radiosurgery for AVMs reduced motor complications without decreasing the obliteration rate\textsuperscript{49}. We are currently testing motor-associated OAR avoidance, along with avoidance of other eloquent white matter tracts, during SRS on a prospective clinical trial in patients with brain metastases\textsuperscript{50}. However, in many instances involving brain tumors, a certain amount of dose to nearby structures is unavoidable. NTCP models may therefore also be used in the future to identify patients at high risk of toxicity, in whom preventative or rehabilitative strategies may be undertaken. The rehabilitation of motor skills has been found to result in persistent gray matter changes over the short term, suggesting this may counteract the atrophy noted in RT and other studies\textsuperscript{51}. The restoration of motor skill after injury may involve enhanced activation of contralateral cortex as well\textsuperscript{52}.

In terms of clinical variables, patient age emerged as the most important demographic/clinical variable. Other studies have confirmed the finding that increasing age may compromise performance on the grooved pegboard test\textsuperscript{4,53,54}, as fine motor dexterity declines with normal and pathological aging. Steroid use emerged as a predictive factor in several of our models for FMS decline. This may be a surrogate marker for tumor volume, edema, mass effect, and/or overall functioning. Prolonged steroid use may also cause myopathy in some patients, though this is usually associated with large muscle groups and not fine motor control. Use of anti-epileptic mediations has also been associated with decline in FMS\textsuperscript{55}, however, this was not a particularly important variable in our analyses.

Our study does have potential limitations. Performance on nested cross-validation was considerably lower than the final models fit to all data, suggesting an unavoidable degree of optimism or overfitting given our modest sample size. Nevertheless, our best models attained acceptable discrimination ability (Random Forest, AUC 0.69) on validation, and the models were well-calibrated. The inclusion of several ROIs and several dosimetric quantities with high
collinearity necessitates variable selection for parsimonious, interpretable models with optimal bias-variance trade-off. Motor ROIs were segmented using robust methods for parcellation of white matter tracts and cortical and subcortical structures; these methods are well-validated in the neuroimaging literature, including in brain tumor patients. To minimize any confounding by tumor, segmentations for each patient were inspected slice by slice, manually censoring tumor, surgical cavities, and edema. The endpoint of decline on fine motor skill tests may be influenced by other biomedical or psychological factors beyond the variables we considered.

Our sample size was modest, though we employed a nested cross-validation approach in order to avoid overfitting as much as possible. The current study sample size is similar or larger than other studies on hippocampal dose-response models. We enrolled a heterogeneous group of brain tumor patients, and tumor type or receipt of chemotherapy were not important variables in our models, though our sample size likely limits these findings. With our heterogeneous group and validation however, our findings are more generalizable to all primary brain tumor patients undergoing intracranial RT. Our models do not exhibit excellent discriminatory power, but perform on par with other recently developed multivariate NTCP models predicting xerostomia, esophagitis, rectal and bladder morbidity, or radiation pneumonitis, for example. Outcomes were prospectively gathered, and objectively measured by a neuropsychologist. Our models rely on a pure structure-function paradigm, which the complexities of the human brain may certainly defy. Network connectivity is a growing area of research utilizing graph theory, borrowed from computer science, where the brain is represented as network of nodes (anatomical areas under consideration) and edges (the interconnections between these areas). Diminished connectivity in motor networks among patients with brain tumors and weakness has been documented, and future studies may consider these more complex networks models to predict motor performance.
In conclusion, we present the first NTCP models for a neurologic functional endpoint outside of hippocampal dose and cognition. Specifically, we found that the dose to several supratentorial motor-associated ROIs correlated with decline on pegboard dominant hand tests at 6 months post-RT in primary brain tumor patients undergoing partial fractionated brain RT. Future studies may externally validate these models and employ prospective strategies to minimize loss of fine motor function after RT.
Figure Captions

**Fig 1.** Heatmap for univariate logistic regression coefficient p-values for decline on PDH

**Fig 2.** A) Frequency of variable selection over the cross-validation folds for the logistic methods.  
B) Variable importance in the random forest model by mean decrease in Gini score.

**Fig 3.** Boxplots of area under the curve (AUC) performance over the repeated outer cross-validation held-out folds.


4. XXXX


18. XXXX

19. XXXX


22. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of


43. Tolle KA, Rahman-Filiapi AM, Hale AC, Kitchen Andren KA, Spencer RJ. Grooved Pegboard Test as a measure of executive functioning. *Appl Neuropsychol.*


50. XXXX


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Pegboard N = 42¹</th>
<th>DKEFS-TM N = 44¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>45.4 (20.0-75.0)</td>
<td>46.2 (20.0-75.0)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25 (60%)</td>
<td>25 (57%)</td>
</tr>
<tr>
<td>Female</td>
<td>17 (40%)</td>
<td>19 (43%)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>39 (92.9%)</td>
<td>41 (93.2%)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>Black</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>3 (7.1%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>39 (93%)</td>
<td>41 (93%)</td>
</tr>
<tr>
<td>Handedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>37 (88%)</td>
<td>39 (89%)</td>
</tr>
<tr>
<td>Left</td>
<td>5 (12%)</td>
<td>5 (11%)</td>
</tr>
<tr>
<td>KPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90-100</td>
<td>39 (93%)</td>
<td>41 (93%)</td>
</tr>
<tr>
<td>&lt;=80</td>
<td>3 (7.1%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Radiation Modality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMRT/VMAT</td>
<td>28 (67%)</td>
<td>30 (68%)</td>
</tr>
<tr>
<td>Proton</td>
<td>14 (33%)</td>
<td>14 (32%)</td>
</tr>
<tr>
<td>Radiation Dose (Gy)</td>
<td>57.0 (50.4-70.0)</td>
<td>57.0 (50.4-70.0)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal</td>
<td>12 (29%)</td>
<td>13 (30%)</td>
</tr>
<tr>
<td>Temporal</td>
<td>12 (29%)</td>
<td>11 (25%)</td>
</tr>
<tr>
<td>Suprasellar</td>
<td>7 (17%)</td>
<td>7 (16%)</td>
</tr>
<tr>
<td>Parietal</td>
<td>4 (9.5%)</td>
<td>5 (11%)</td>
</tr>
<tr>
<td>Cerebellar</td>
<td>2 (4.8%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Base of Skull</td>
<td>2 (4.8%)</td>
<td>2 (4.5%)</td>
</tr>
<tr>
<td>Cavernous Sinus</td>
<td>2 (4.8%)</td>
<td>2 (4.5%)</td>
</tr>
<tr>
<td>Sphenoid Wing</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>Laterality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>22 (52%)</td>
<td>21 (48%)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Pegboard N = 42</td>
<td>DKEFS-TM N = 44</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Right</td>
<td>16 (38%)</td>
<td>19 (43%)</td>
</tr>
<tr>
<td>Central</td>
<td>4 (9.5%)</td>
<td>4 (9.1%)</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glioma: High Grade</td>
<td>17 (40%)</td>
<td>18 (41%)</td>
</tr>
<tr>
<td>Glioma: Low Grade</td>
<td>9 (21%)</td>
<td>9 (20%)</td>
</tr>
<tr>
<td>Meningioma</td>
<td>9 (21%)</td>
<td>10 (23%)</td>
</tr>
<tr>
<td>Pituitary Adenoma</td>
<td>3 (7.1%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Craniopharyngioma</td>
<td>2 (4.8%)</td>
<td>2 (4.5%)</td>
</tr>
<tr>
<td>Chondrosarcoma</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>Schwannoma</td>
<td>1 (2.4%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td><strong>Surgery</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTR</td>
<td>9 (21%)</td>
<td>10 (23%)</td>
</tr>
<tr>
<td>STR</td>
<td>27 (64%)</td>
<td>27 (61%)</td>
</tr>
<tr>
<td>Biopsy</td>
<td>2 (4.8%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>None</td>
<td>4 (9.5%)</td>
<td>4 (9.1%)</td>
</tr>
<tr>
<td><strong>Steroid Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 (40%)</td>
<td>17 (39%)</td>
<td></td>
</tr>
<tr>
<td><strong>Seizures</strong></td>
<td>17 (40%)</td>
<td>18 (41%)</td>
</tr>
<tr>
<td><strong>Antiepileptic Drug Use</strong></td>
<td>23 (55%)</td>
<td>24 (55%)</td>
</tr>
<tr>
<td><strong>Concurrent Chemotherapy</strong></td>
<td>20 (48%)</td>
<td>21 (48%)</td>
</tr>
<tr>
<td><strong>Adjuvant Chemotherapy</strong></td>
<td>25 (60%)</td>
<td>26 (59%)</td>
</tr>
<tr>
<td><strong>Progression at 6mo</strong></td>
<td>3 (7.1%)</td>
<td>3 (6.8%)</td>
</tr>
</tbody>
</table>

1 N (%) or Median (range)
<table>
<thead>
<tr>
<th>Cerebellum WM</th>
<th>Cerebellum Cortex</th>
<th>Paracentral WM</th>
<th>Decision Stump</th>
<th>Logistic Regression</th>
<th>Spearman Correlation</th>
<th>Non-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff (Gy)</td>
<td>AUC (95% CI)</td>
<td>OR (95% CI)</td>
<td>p-value</td>
<td>Rs</td>
<td>p-value</td>
<td></td>
</tr>
</tbody>
</table>

**Corticospinal Tract**
- $D_{\text{max}}$: 53.54, 0.71 (0.57 - 0.85)
- $D_{\text{mean}}$: 19.53, 0.73 (0.59 - 0.87)

**Prefrontal Cortex**
- $D_{\text{max}}$: 24.27, 0.76 (0.62 - 0.89)
- $D_{\text{mean}}$: 13.47, 0.74 (0.60 - 0.88)

**Prefrontal WM**
- $D_{\text{max}}$: 27.09, 0.75 (0.61 - 0.89)
- $D_{\text{mean}}$: 15.40, 0.74 (0.60 - 0.88)

**Postcentral Cortex**
- $D_{\text{max}}$: 43.80, 0.71 (0.57 - 0.85)
- $D_{\text{mean}}$: 17.70, 0.70 (0.56 - 0.84)

**Postcentral WM**
- $D_{\text{max}}$: 44.83, 0.73 (0.59 - 0.87)
- $D_{\text{mean}}$: 12.09, 0.72 (0.58 - 0.86)

**Paracentral Cortex**
- $D_{\text{max}}$: 6.71, 0.72 (0.57 - 0.86)
- $D_{\text{mean}}$: 3.84, 0.69 (0.54 - 0.83)

**Paracentral WM**
- $D_{\text{max}}$: 59.46, 0.59 (0.49 - 0.68)
- $D_{\text{mean}}$: 3.59, 0.67 (0.52 - 0.81)

**Caudate**
- $D_{\text{max}}$: 58.88, 0.69 (0.56 - 0.83)
- $D_{\text{mean}}$: 33.70, 0.76 (0.63 - 0.90)

**Pallidum**
- $D_{\text{max}}$: 38.13, 0.72 (0.58 - 0.86)
- $D_{\text{mean}}$: 20.24, 0.71 (0.56 - 0.85)

**Putamen**
- $D_{\text{max}}$: 40.94, 0.72 (0.58 - 0.86)
- $D_{\text{mean}}$: 22.40, 0.73 (0.59 - 0.87)

**Thalamus**
- $D_{\text{max}}$: 53.36, 0.74 (0.60 - 0.88)
- $D_{\text{mean}}$: 27.31, 0.77 (0.63 - 0.90)

**Cerebellum Cortex**
- $D_{\text{max}}$: 43.12, 0.64 (0.52 - 0.75)
- $D_{\text{mean}}$: 17.54, 0.62 (0.51 - 0.73)

**Cerebellum WM**
- $D_{\text{max}}$: 41.88, 0.64 (0.52 - 0.75)
- $D_{\text{mean}}$: 23.30, 0.62 (0.51 - 0.73)

**Demographic/Clinical**
- Adjuvant Chemotherapy: 0.59 (0.44 - 0.74)
- Age: 60.50, 0.64 (0.51 - 0.77)
- Anti-Epileptic Drug Use: 0.63 (0.48 - 0.78)
- Concurrent Chemotherapy: 0.54 (0.39 - 0.70)
- Diagnosis (Other): 0.57 (0.42 - 0.72)
- Dose: 60.60, 0.56 (0.48 - 0.64)
- Ethnicity (Non-Hispanic): 0.56 (0.49 - 0.63)
- Handedness (R): 0.50 (0.40 - 0.60)

p-values for significance of adj. coeff: 0.01
Table 2. Univariate Analyses for Dosimetric and Demographic/Clinical variables and association with decline on PDH at 6 months.

<table>
<thead>
<tr>
<th>Variable</th>
<th>KPS (≤ 80)</th>
<th>Laterality (Dominant)</th>
<th>Progression at 6mo</th>
<th>Race (Other)</th>
<th>Radiation Modality (Protons)</th>
<th>Seizures</th>
<th>Sex (M)</th>
<th>Steroid Use</th>
<th>Surgery (any)</th>
<th>Surgery (GTR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.54 (0.45 - 0.63)</td>
<td>3.20 (0.28 - 72.45)</td>
<td>0.36</td>
<td>0.15</td>
<td>0.35</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.68 (0.54 - 0.83)</td>
<td>4.87 (1.31 - 21.49)</td>
<td>0.02</td>
<td>0.36</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51 (0.43 - 0.59)</td>
<td>0.72 (0.03 - 8.13)</td>
<td>0.79</td>
<td>-0.04</td>
<td>0.80</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.53 (0.44 - 0.62)</td>
<td>0.46 (0.02 - 3.96)</td>
<td>0.52</td>
<td>-0.10</td>
<td>0.52</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.52 (0.37 - 0.67)</td>
<td>1.16 (0.31 - 4.28)</td>
<td>0.82</td>
<td>0.03</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65 (0.50 - 0.80)</td>
<td>3.67 (1.03 - 14.24)</td>
<td>0.05</td>
<td>0.31</td>
<td>0.05</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51 (0.35 - 0.66)</td>
<td>0.95 (0.27 - 3.41)</td>
<td>0.94</td>
<td>-0.01</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.70 (0.56 - 0.85)</td>
<td>5.81 (1.57 - 24.21)</td>
<td>0.01</td>
<td>0.41</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.52 (0.41 - 0.63)</td>
<td>0.70 (0.09 - 4.08)</td>
<td>0.70</td>
<td>-0.06</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.52 (0.39 - 0.65)</td>
<td>0.81 (0.18 - 3.82)</td>
<td>0.79</td>
<td>-0.04</td>
<td>0.79</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Logistic</td>
<td>LASSO</td>
<td>LASSO (w/ FS)</td>
<td>Random Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------</td>
<td>---------</td>
<td>---------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Final Fitted Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.75</td>
<td>0.79</td>
<td>0.80</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balanced Accuracy</td>
<td>0.71</td>
<td>0.71</td>
<td>0.76</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brier Score</td>
<td>0.19</td>
<td>0.20</td>
<td>0.18</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration Slope</td>
<td>1.00</td>
<td>2.89</td>
<td>1.39</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration Intercept</td>
<td>0.00</td>
<td>0.59</td>
<td>0.11</td>
<td>-0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosmer–Lemeshow p-value</td>
<td>0.67</td>
<td>0.07</td>
<td>0.23</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.25</td>
<td>0.22</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                      |          |         |               |               |
| **Nested Cross Validation** |      |         |               |               |
| AUC                  | 0.61 (SD 0.005) | 0.55 (SD 0.005) | 0.63 (SD 0.005) | 0.69 (SD 0.006) |
| Balanced Accuracy    | 0.62 (SD 0.005) | 0.57 (SD 0.003) | 0.62 (SD 0.004) | 0.67 (SD 0.005) |
| Brier Score          | 0.26 (SD 0.002) | 0.26 (SD 0.002) | 0.24 (SD 0.002) | 0.23 (SD 0.002) |

**Table 3.** Performance and characteristics for each model-building method.