

## Unsupervised machine learning models reveal predictive clinical markers of glioblastoma patient survival using white blood cell counts prior to initiating chemoradiation

Wesley Wang<sup>○</sup>, Zeynep Temerit Kumm, Cindy Ho, Ideli Zanesco-Fontes, Gustavo Texiera, Rui Manuel Reis<sup>○</sup>, Horacio Martinetto, Javaria Khan, Martin G. McCandless, Katherine E. Baker, Mark D. Anderson, Muhammad Omar Chohan, Sasha Beyer, J. Brad Elder, Pierre Giglio<sup>○</sup>, and José Javier Otero<sup>○</sup>

*Department of Pathology, Ohio State University Wexner Medical Center, Columbus, Ohio, USA (W.W., Z.T.K., C.H., J.J.O.); Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil (I.Z.-F., R.M.R.); Department of Pathology, Barretos Cancer Hospital, Barretos, Brazil (G.T.); Life and Health Sciences Research Institute (ICVS)/School of Medicine, University of Minho, Braga, Portugal (R.M.R.); ICVS/3B's—PT Government Associate Laboratory, Braga-Guimarães, Portugal (R.M.R.); Departamento de Neuropatología y Biología Molecular, Instituto de Investigaciones Neurológicas Dr Raúl Carrea (FLENI), Buenos Aires, Argentina (H.M.); Department of Pathology, University of Mississippi Medical Center, Jackson, Mississippi, USA (J.K.); Department of Neuro-Oncology, University of Mississippi Medical Center, Jackson, Mississippi, USA (M.G.MC., K.E.B., M.D.A.); Department of Neurosurgery, University of Mississippi Medical Center, Jackson, Mississippi, USA (M.G.MC., K.E.B., M.O.C.); Department of Radiation Oncology, Ohio State University Wexner Medical Center, Columbus, Ohio, USA (S.B.); Department of Neurosurgery, Ohio State University Wexner Medical Center, Columbus, Ohio, USA (J.B.E.); Department of Neurology, Ohio State University Wexner Medical Center, Columbus, Ohio, USA (P.G.)*

Corresponding Authors: José Javier Otero, MD, PhD, Department of Pathology, Ohio State University College of Medicine, 4166 Graves Hall, 333 W 10th Avenue Columbus, OH 43210, USA ([jose.otero@osumc.edu](mailto:jose.otero@osumc.edu)); Pierre Giglio, MD, Department of Neurology, Ohio State University Wexner Medical Center, M410 Starling Loving Hall, 320 W 10th Avenue Columbus, OH 43210, USA ([pierre.giglio@osumc.edu](mailto:pierre.giglio@osumc.edu)).

### Abstract

**Background.** Glioblastoma is a malignant brain tumor requiring careful clinical monitoring even after primary management. Personalized medicine has suggested the use of various molecular biomarkers as predictors of patient prognosis or factors utilized for clinical decision-making. However, the accessibility of such molecular testing poses a constraint for various institutes requiring identification of low-cost predictive biomarkers to ensure equitable care.

**Methods.** We collected retrospective data from patients seen at Ohio State University, University of Mississippi, Barretos Cancer Hospital (Brazil), and FLENI (Argentina) who were managed for glioblastoma—amounting to 581 patient records documented using REDCap. Patients were evaluated using an unsupervised machine learning approach comprised of dimensionality reduction and eigenvector analysis to visualize the inter-relationship of collected clinical features.

**Results.** We discovered that the serum white blood cell (WBC) count of a patient during baseline planning for treatment was predictive of overall survival with an over 6-month median survival difference between the upper and lower quartiles of WBC count. By utilizing an objective PD-L1 immunohistochemistry quantification algorithm, we were further able to identify an increase in PD-L1 expression in glioblastoma patients with high serum WBC counts.

**Conclusions.** These findings suggest that in a subset of glioblastoma patients the incorporation of WBC count and PD-L1 expression in the brain tumor biopsy as simple biomarkers predicting glioblastoma patient survival. Moreover, machine learning models allow the distillation of complex clinical data sets to uncover novel and meaningful clinical relationships.

### Key Points

- Unsupervised learning models visualize complex data sets to uncover novel relationships.
- White blood cell count and PD-L1 expression are simple biomarkers predictive of glioblastoma patient survival.

### Importance of the Study

Glioblastomas remain a challenging malignancy to manage following surgical resection and chemoradiation. Efforts to identify prognostic markers of disease are therefore critical in ascertaining how a provider will manage a patient's tumor. Current efforts have largely focused on molecular markers to stratify patient populations, but these approaches are not equally accessible. Building upon studies that have evaluated routinely collected clinical data, we highlight

the use of machine learning-based approaches to rapidly visualize relationships in clinical data to uncover novel trends. Moreover, the use of pretreatment blood count may play a role in patient prognostication and tumor microenvironment status. These findings thus highlight the importance of future exploration of patient tumor-immune activity while lowering the complexity of assessment through machine learning.

Glioblastoma (GB) patients suffer from aggressive solid tumors of the central nervous system, with 95% projected to be deceased within 5 years following diagnosis.<sup>1</sup> Cause of death in glioblastoma varies but may include herniation secondary to mass effect, treatment complications, and aspiration pneumonia due to brainstem dysfunction.<sup>2</sup> Despite aggressive adjuvant therapy, however, almost all patients will experience tumor progression. To address this, major strides have been made within neuro-oncology to merge personalized medicine to better predict patient outcomes. Although therapies for GB management after maximal-safe surgical resection have remained largely unchanged since the concomitant use of temozolomide (TMZ) with radiation (ChemoRT), molecular biomarkers have been integrated as critical tools in the diagnosis and prognostication of gliomas.<sup>3,4</sup> From a diagnostic standpoint, the lack of *IDH* mutation defines the current World Health Organization (WHO) 2021 definition of GB.<sup>4</sup> Moreover, *TERT* promoter mutations, EGFR amplifications, and modifications of chromosome 7/9/10 are all recognized molecular changes present in GB.

Prognostically, several molecular markers have been suggestive of predicting outcomes in GB patients. Namely, *MGMT* promoter methylation status is routinely tested to predict TMZ efficacy due to the antagonistic role of MGMT-mediated repair following TMZ-induced DNA alkylation.<sup>5,6</sup> However, other markers such as *CDKN2A/B* loss and *EGFRvIII* have been shown in retrospective studies to act as poor prognostic markers in GB patients or subset populations.<sup>7,8</sup> The need for prognostic markers of disease is critical in ascertaining whether certain patients should be monitored more closely during follow-up. Furthermore, gold standard management of GB requires aggressive multimodal treatment amongst neuro-oncology, neurosurgery, neuroradiology, and radiation oncology which

would largely benefit from improved triaging methods that prioritize which patients should or should not have such care. Moreso, precise prognostic grading benefits patients and families when discussing management direction.

Unfortunately, molecular marker accessibility is not equal across healthcare ecosystems.<sup>9,10</sup> Current discussions in the field of neuro-oncology have pointed to major accessibility barriers and bioethical implications of pure reliance on molecular biomarkers of disease to understand cancer.<sup>10</sup> Molecular pathology testing requires batching to reduce patient costs, resulting in significant turn-around time delays for molecular assays.<sup>11</sup> In consequence, groups have sought ways to evaluate outcomes of glioma patients using surrogate measures such as neurocognitive testing, psychiatric examination, or image analysis of histology to predict molecular phenotypes as these turn-around-times are superior to those of molecular pathology.<sup>9,12,13</sup> Although promising, however, integration of these approaches will take time whereas routinely collected data is readily available.

To better explore potential prognostic markers already routinely collected while managing GB patients, we retrospectively evaluated 581 patients at 4 sites from 3 countries: Ohio State University (USA), University of Mississippi (USA), Barretos Cancer Hospital (Brazil), and FLENI (Argentina). Patient features were assessed using an unsupervised learning approach to understand how clinical metrics related to each other—including relevant clinical endpoints such as progression-free survival (PFS) and overall survival (OS). Furthermore, by utilizing these novel relationships, we better characterized the relevance of routine complete blood counts (CBCs) and merged its utility with IHC staining in pathology to suggest novel workflows that may predict outcomes of patients with GB.

## Materials and Methods

### Selection Criteria and Clinical Data Collection

Evaluation of retrospective clinical charts from the electronic health record (EHR) was performed at the Ohio State University (OSU) with IRB approval under study number 2020C0062, University of Mississippi (UMMC) with IRB approval under study number UMMC-IRB-2022-93, Barretos Cancer Hospital with IRB approval under study number 1604/2018, and FLENI under ethics committee/patient's informed consent approval. Clinical records from patients receiving GB care from 2012 to 2020 were evaluated. Inclusion criteria for patients were designated by prior history of GB treatment with total surgical resection and ChemoRT. Following WHO 2021 guidelines, assessed GBs were defined using Grade-4 pathology and confirmation of *IDH* wild-type status.<sup>4</sup> Prior histories of low-grade gliomas, *IDH* mutation, or patients with poor documentation of disease course were excluded. Study data were collected and managed using REDCap electronic data capture tools.<sup>14,15</sup> REDCap was utilized to (a) securely store and deidentify patient records for downstream use, (b) ensure consistency of data collection, and (c) be distributable to collaborators desiring to replicate or collect similar metrics as described in the study (Supplementary Document 1).

Clinical features were collected as listed in the EHR. Comorbidity scoring was performed by manually listing known comorbidities and scoring them in REDCap following the Charlson Comorbidity Index (CCI).<sup>16</sup> Karnofsky performance scale (KPS) was recorded at the time of initial diagnosis to assess patient functional status. Lesion and molecular features were extracted from radiology and pathology reports, respectively. Results from CBCs—including white blood cell count (WBC), neutrophil count, lymphocyte count, and platelet count were designated as CBC draws occurring approximately 2–4 weeks after surgery during patient follow-up with neuro-oncology prior to beginning ChemoRT. The neutrophil-to-lymphocyte ratio (NLR) was calculated by dividing the neutrophil count by the lymphocyte count in a patient. Steroid dose was defined as the total daily dexamethasone steroid intake during the same day as CBC collection. Variables containing dates were converted to deidentified values of time to remove potential patient timeline identification after calculating relevant timespans in days. OS in the study was defined as the time from primary tumor resection until the time of death. PFS was defined as time from primary tumor resection until the time of initial detection of a novel enhancing lesion on imaging. Confirmation of enhancement as being either cancer-recurrent or treatment-reactive in nature was completed by clinical correlation and consensus from a multidisciplinary tumor board.

### Exploration of Clinical Features

Collected clinical data from REDCap was exported and deidentified for use in R. Missingness of data was

evaluated using the *nanjar* package.<sup>17</sup> Cases with over 30% missing records were not evaluated using unsupervised analysis while any missing features were imputed using the multiple imputation by chained equations (*mice*) package in R in order to perform principal component analysis (PCA) visualization alone (Supplementary Figure 1).<sup>18</sup> Using the *factoextra* package, PCA was performed over the data set and relations of clinical features were visualized using eigenvector plotting of our PCA as described in our previous work.<sup>19,20</sup>

Returning to our non-imputed data, the exploration of clinical features as a function of OS was explored using Cox regression modeling with a specific assessment of the relationships uncovered in our PCA visualization. Clinical features were independently evaluated using a univariate Cox model to assess the prediction of OS time. *P*-values less than .05 were omitted in subsequent modeling. Univariate-significant features were then evaluated using a multivariate Cox regression model using patient records found to have no missing relevant features. Patient survival was further explored by quartile stratification of patients for relevant clinical features (Supplementary Table 1). The highest and lowest 25% groups were compared for PFS and OS using the *survival* and *survminer* packages.<sup>21,22</sup> Visualization of relevant clinical confounders was performed using R with *ggplot2*.<sup>23</sup>

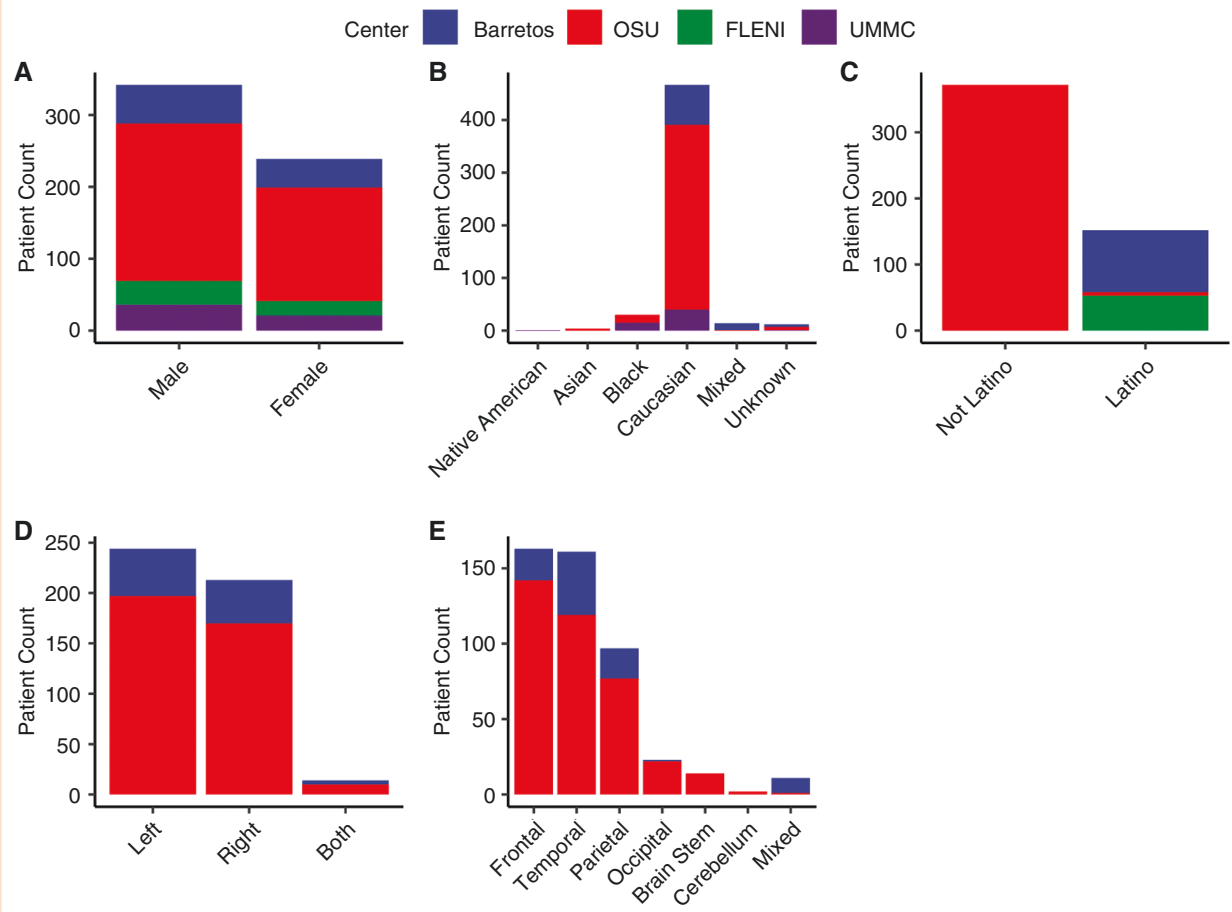
### PD-L1 Image Analysis

PD-L1 immunohistochemistry imaging was collected from relevant patient samples at Ohio State which underwent routine PD-L1 evaluation at the time of primary surgery for GB. Image tiles were collected from digital pathology slides and processed in R using *EImage*.<sup>24</sup> RGB images were deconvoluted for hematoxylin and DAB stain layers using *sci-kit-image* and *reticulate*.<sup>25,26</sup> Segmentation for either hematoxylin or DAB staining was performed using Otsu thresholding.<sup>27</sup> Segmented regions were calculated for morphologic features using *EImage*. Filtration of segmentation was performed using a random-forest-based classifier trained over intensity-based morphology features to classify segmentation as no stain, low stain, medium stain, and high stain (Supplementary Figure 2).<sup>28</sup> No stain segmentations were filtered. PD-L1 staining was represented by the ratio of DAB-stained pixels to hematoxylin-stained pixels—calculated from segmentation areas in *EImage* based upon objective stain scoring approaches delineated by Igarashi et al.<sup>29</sup>

## Results

### Patient Characteristics

A total of 581 patients were evaluated and documented in REDCap across 4 centers. In turn, overall demographic scores and biases across centers were assessed. Overall, the mean age was 61 years (range: 20–89; Supplementary Table 2). Men slightly exceeded woman in representation across centers (Figure 1A). The distribution of ethnic origin is directly related to the site of collection. North



**Figure 1.** Distribution of patient demographics and lesion characteristics with respect to center. Bar plot distributions of (A) gender, (B) race, (C) ethnicity, (D) lesion sidedness, and (E) lesion lobe location across collected centers. Centers are represented as OSU (red), UMMC (purple), Barretos (blue), and FLENI (green). Omission of center in plot indicates variable was not collected.

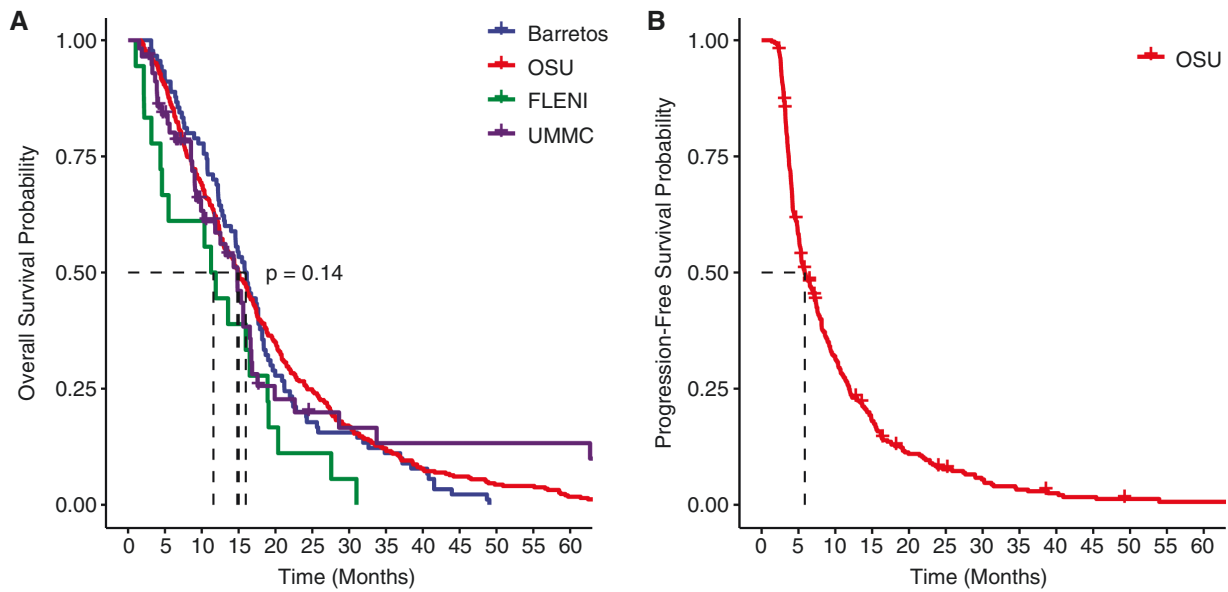
American sites represented predominately Caucasians and Black/African Americans with non-Latino ethnicity, while South American sites were of Caucasian and mixed ancestry—that in U.S. surveys would be considered Hispanic/Latino (Figure 1B and C). All patients had biopsy confirmed GB with total resection of tumor at primary surgery. There was a slight predominance of left-sided lesions primarily occurring in the frontal, temporal, and parietal lobes across sites (Figure 1D and E). On average, tumor diameter was 4.34 cm (range: 0.1–9.50 with most without midline shift on imaging with KPS near 80 (range: 30–100; Supplementary Table 2). It was however noted that several clinical features were not collected in sites outside OSU (Supplementary Figure 1). When tested, IHC studies showed primarily ATRX locus intact (98%) and p53 mutation (78%) as defined by positivity in over 10% of cells. Predominant molecular features were *EGFR* amplification (56%) and unmethylated *MGMT* promoter status (57%).

ChemoRT overall followed a traditional 60 Gray-30 fractions radiation plan, but some patients received a hypo-fractionated regimen or did not complete ChemoRT (Radiation Plan: 55.59 [range: 5.34–75]; Radiation

Fractions: 26.8 [range: 1–50]). Adjuvant TMZ therapy was completed at 3 cycles on average (range: 0–19). Symptomatic management of edema with steroids at the time of CBC collection was on average 2.84 mg/day with a broad range of use (range: 0–24). Outcomes of patients were assessed as both PFS and OS from the date of primary surgery. Among centers, the median OS ranged from 12 to 16 months without significant difference among groups while the median PFS was 6 months at OSU (Figure 2).

### Unsupervised Analysis of Patient Reveals Capabilities of CBCs in Predicting Overall Survival Outcome and Time to Enhancement

Utilizing the various collected clinical data points, we sought to evaluate whether specific clinical features were correlated with relevant outcomes in patients that have not been previously integrated into clinical practice. Specifically, we posited that applying an unsupervised machine learning approach could reveal clinical relationships amongst features permitting us to visualize novel findings predictive in GB patient prognosis. To do so, our



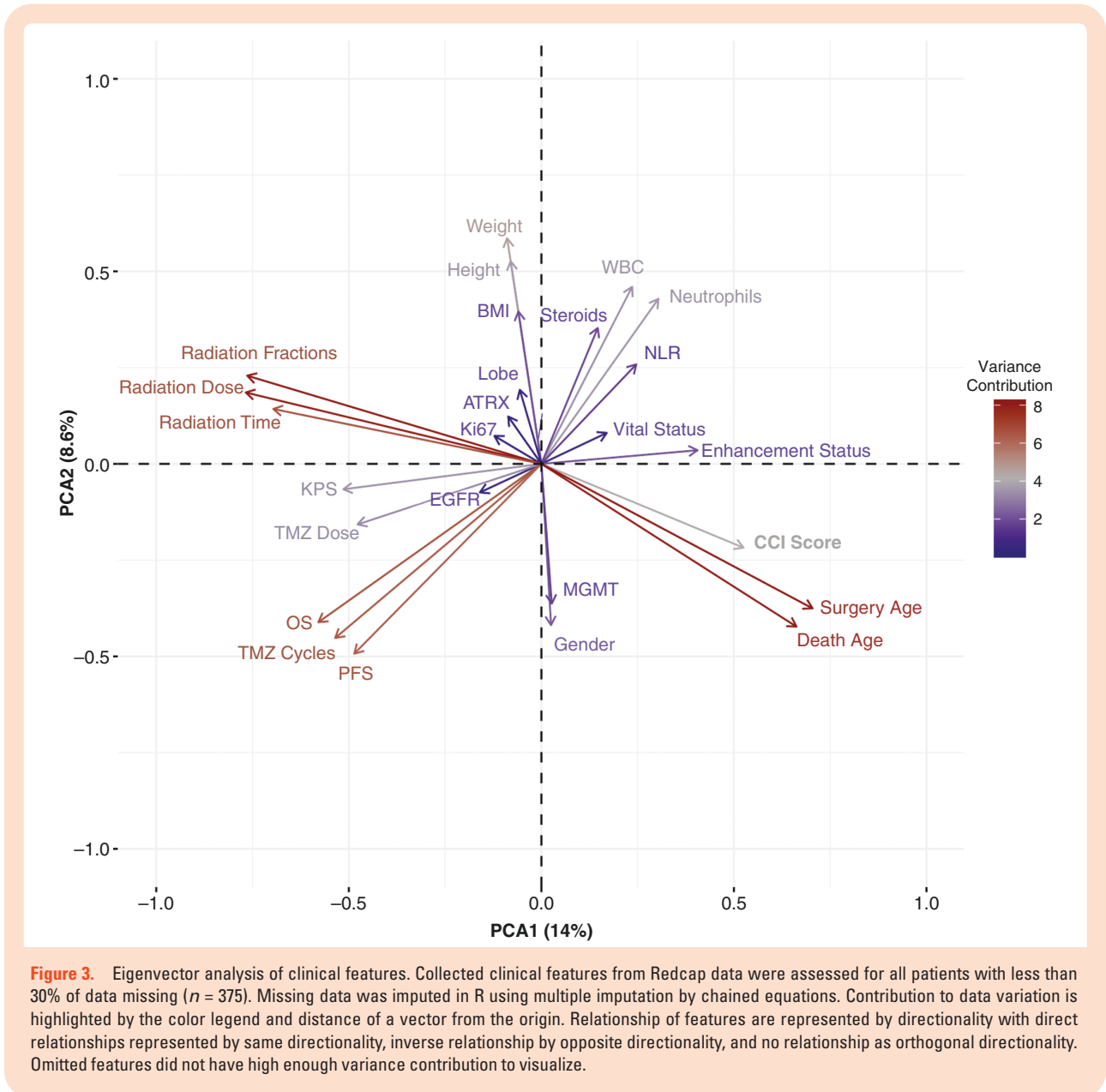
**Figure 2.** Overall survival and progression-free survival across sites. (A) OS and (B) PFS of patients across sites. Centers are represented as OSU (red), UMMC (purple), Barretos (blue), and FLENI (green).

multidimensional data set was reduced using principal components analysis (PCA) to stratify patients by these clinical metrics and visualized using PCA eigenvector plotting (Figure 3). The directionality of eigenvectors (arrows) amongst other eigenvectors represents direct correlations through same arrow directionality, inverse correlations through opposite directionality, and no correlation through orthogonal directionality. With these considerations, we identified relationships of features to relevant clinical outcomes. Notably, the directionality of eigenvectors for OS and PFS occurred similarly [lower left quadrant] with inverse directionality to CBC-related measures—note that WBC and neutrophil measurements are in the upper right quadrant indicating that as WBC increased OS and PFS decreased (Figure 3). In contrast, enhancement status was shown on the right side of the plot, but other vectors had less robust directional relationships—with the most prominent being inverse directionality of KPS and adjuvant TMZ dosage in the left quadrant. Nevertheless, as our eigenvector plot uncovered novel variations amongst clinical features, we further explored these findings using regression modeling.

Cox regression modeling was applied to predict OS based on our clinical features. Based on our PCA eigenvector plot, we theorized that CBC-related metrics would be significant in predicting survival time. Univariate models were first performed over the study population to assess which features were found to have significance. In total, 12 separate features were found to be significant (Patient Age, KPS, CCI score, *MGMT* methylation status, WBC count, Neutrophil count, NLR, Radiation Dose, Radiation Fractions, Overall Radiation Time, Adjuvant TMZ Cycles, and Adjuvant TMZ Dose; Supplementary Table 3). A multivariate Cox regression model was constructed using 9 significant features from Supplementary Table 3 which were

not derived from each other (ie, neutrophil load is represented within WBC load or overall radiation time is affected by the fractionation of radiation). Six clinical features were found to be relevant (Patient Age at surgery, KPS at diagnosis, *MGMT* methylation status, WBC count at follow-up, Total Radiation Dose, and Completed Adjuvant TMZ Cycles; Supplementary Table 4). Although the contribution of WBC count was significant in our multivariate model, the hazard ratio (HR) was small (1.023 [1.001–1.046]). In consequence, although a small, but significant risk to poorer survival was evidenced by increased WBC load, we next sought to explore the causes of the observed difference our Cox modeling showed against our PCA eigenvector plot.

Although Cox regression modeling validated the finding that CBCs have predictive capabilities for survival time, we further examined the discordance of our strong correlations seen in the PCA eigenvector plot against the smaller HRs calculated in our Cox models. Specifically, we hypothesized the differences seen in our results may be underscored by robust survival differences present in the extremes of our CBC metrics. Specifically, based on our univariate Cox results, we explored survival differences in patients when stratified by WBC count, neutrophil count, and neutrophil:lymphocyte ratio. CBC measures were evaluated by stratifying populations into quartiles with the lower 25% (Lo in blue) and upper 25% (hi in red) of patients evaluated. In both overall WBC load and neutrophil load, PFS was significantly worse in the Hi group relative to the Lo group ( $P = .0082$  and  $P = .039$ , respectively; Figure 4A and B). Furthermore, evaluation of OS using WBC load and neutrophils showed similar trends between groups ( $P = .00042$  and  $P = .0007$ , respectively). However, while NLR did show a significant survival difference in OS ( $P = .0081$ ), the difference in PFS between groups did not reach our threshold of significance (Figure 4C). Combined

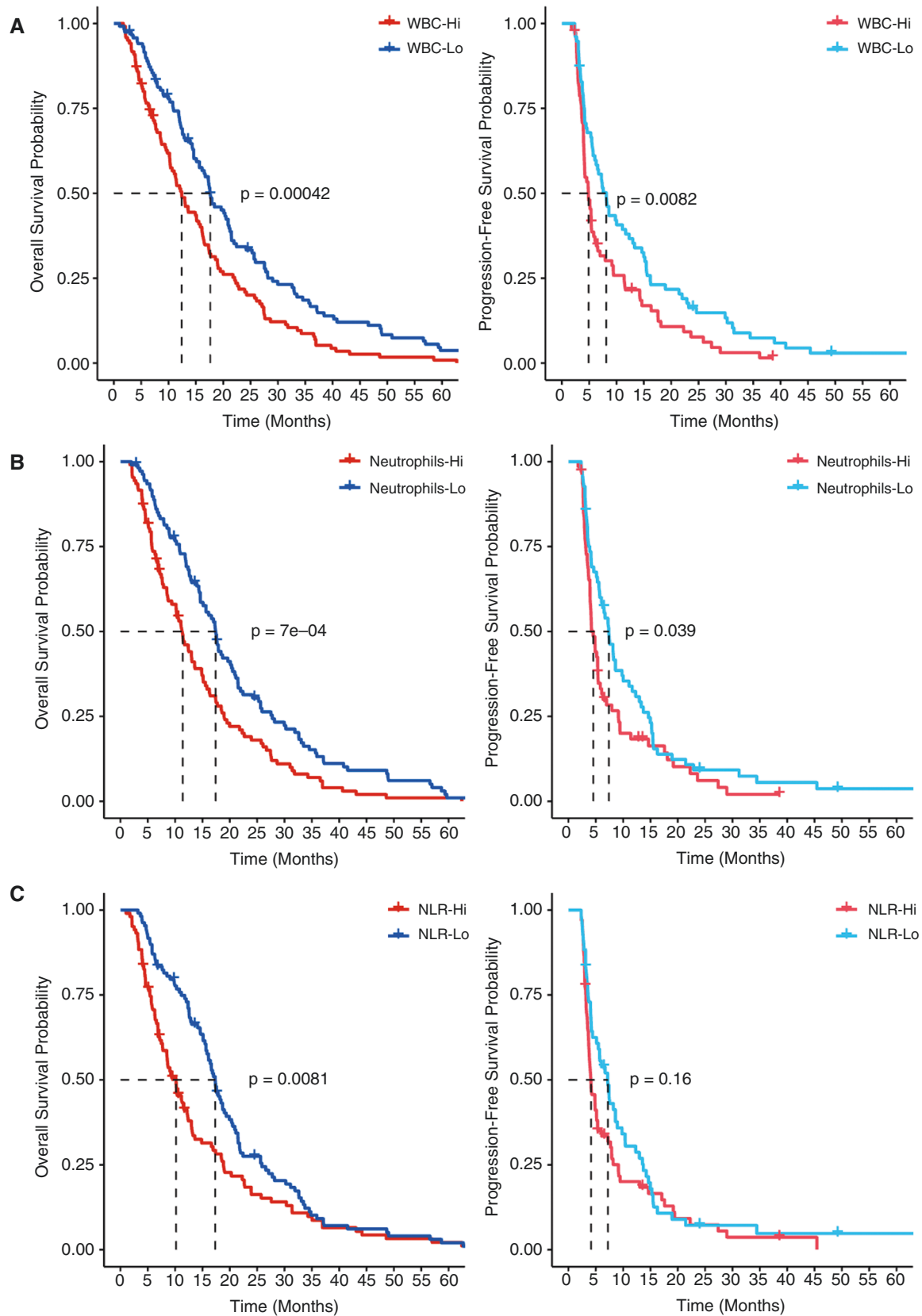


with the previous analyses, our findings strongly support the use of CBCs in predicting both OS and PFS in patients. Specifically, the evaluation of routinely drawn WBC and neutrophil count at the time of pre-ChemoRT planning is shown to be correlative with poorer survival outcomes in patients with high load compared to patients with low load.

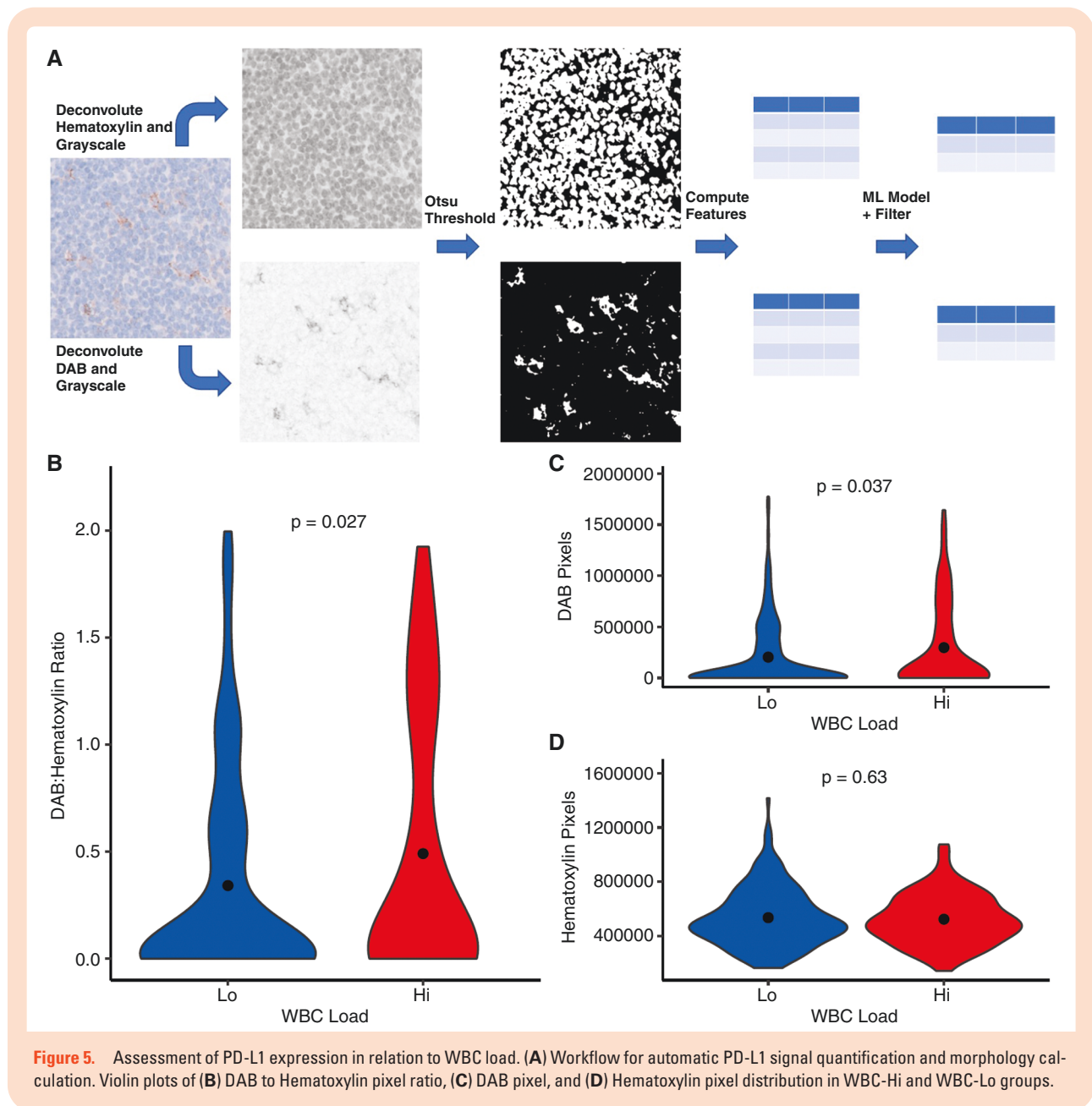
### WBC Load is Reflective of Intrinsic Tumor Microenvironment Changes Present in Glioblastoma

As the measures of WBC load evaluate circulating immune counts, we posited these differences in peripheral immune activity may correlate with intrinsic tumor microenvironment differences found in primary GB events.

To assess this, we selected PD-L1 IHC staining done in a subset of patients during clinical evaluation. Amongst our study population, 57 cases had been evaluated by neuropathology for PD-L1 expression, and representative images were collected from cases and segmented using computer vision techniques to objectively quantify staining (Figure 5A). Staining was quantified as the ratio of DAB-positive PD-L1 staining against hematoxylin nuclear staining to control for tissue cellularity. In turn, increased detection of PD-L1 staining is represented by an increased DAB:hematoxylin ratio. It was observed that the WBC-Hi group showed higher ratios of PD-L1 DAB to hematoxylin staining when compared to WBC-Lo ( $P = .027$ , Figure 5B). Furthermore, assessing the 2 sides of the ratio comparison it was seen that while the amount of DAB pixels detected in an image was higher in the WBC-Hi group ( $P = .037$ ), the detection of hematoxylin pixels did not vary between



**Figure 4.** KM curve survival outcomes based upon CBCs. OS (left) and PFS (right) of patients in upper 25% [red] and lower 25% [blue] of cases stratified based on **(A)** WBC count ( $n = 186/\text{group}$ ), **(B)** Neutrophil count ( $n = 212/\text{group}$ ), and **(C)** Neutrophil-to-Lymphocyte ratio ( $n = 217/\text{group}$ ). All CBCs were collected prior to initiating ChemoRT as routine baseline by neuro-oncology.



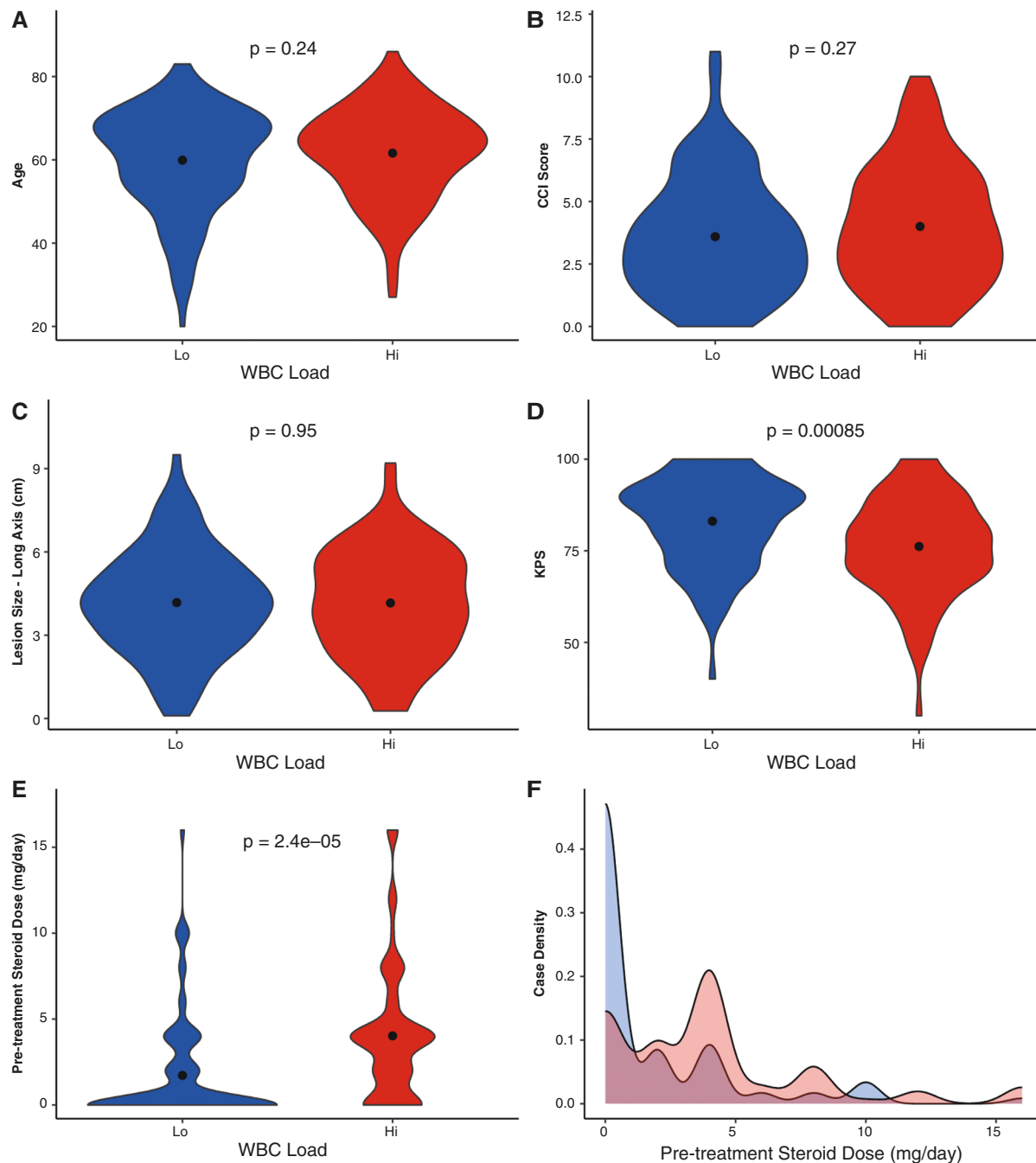
groups ( $P=0.63$ ; [Figure 5C](#) and [D](#)). In conclusion, the increased PD-L1 staining ratio in the WBC-Hi group was not a product of increased cellularity as the distribution of hematoxylin was not different. Overall, these findings indicate that an increase in PD-L1 staining was correlated with the WBC-Hi group which showed poorer survival outcomes in patients—validating that the observed differences in WBC load are correlative to initial immune activity present in GB lesions at resection.

### Steroid Tapering is Highly Heterogenous Following Surgery and May Influence WBC Load

Although our analyses found a strong correlation between CBC load to survival outcomes, we further

evaluated potential clinical confounders that may influence CBC levels prior to ChemoRT. As shown in the analyses of our study population, heterogeneity in patient demographics, lesion characteristics, and patient management was present ([Supplementary Table 2](#)). In turn, an assessment to identify whether specific clinical features significantly varied between our Hi and Lo populations was critical. As patient age, CCI score, KPS, and *MGMT* methylation status were predictive of survival in our univariate Cox model, we assessed these factors in addition to other clinical features that were shown to be correlative to CBC measures in the eigenvector plot (lesion size and steroid intake; [Figure 3](#) and [Supplementary Table 3](#)). Comparing WBC-Hi and Lo groups, *MGMT* methylation status distribution was not found to significantly vary between groups ( $\chi^2=0.88$ ,  $P$ -value = .35).





**Figure 6.** Functional status and steroid tapering vary amongst WBC-Hi/Lo groups. Groups were compared between (A) patient age, (B) CCI score, (C) lesion size, (D) KPS, and (E) steroid dosing at time of CBC collection using violin plots. (F) Distribution of daily dexamethasone steroid intake is visualized using a density plot between groups.

However, while patient age, CCI score, and lesion size were found to not vary between groups, significant variations in KPS ( $P = 8.5e-04$ ) and steroid dosing ( $P = 2.4e-05$ ) between groups were present with the WBC-Hi group showing a higher mean daily steroid intake compared to those in the Lo group (Figure 6A-E). Evaluating the distribution of steroid doses given to patients at the time of post-surgical CBC, a larger percentage of the patient

from the WBC-Lo fully tapered off steroids (Hi: 24.2%; Lo: 65.1%). Nevertheless, an assay of patient distribution shows 55.9% of patients across both groups remain on steroids at follow-up prior to initiating ChemoRT (Figure 6 F). Assessment of survival between Hi and Lo groups in patients that fully tapered off steroids showed significant PFS differences when stratified by WBC load ( $P = .045$ ) and neutrophil load ( $P = .026$ ), but significant trends were

not seen for OS (Supplementary Figure 3), Overall, these findings not only highlight the additional confounding role steroid intake and patient functional status may have on the survival differences seen between CBC load, but also emphasizes the heterogeneity of steroid dosing in patients following GB resection.

## Discussion

### PCA Eigenvector Visualization can Better Uncover Clinical Relationships in Complex Clinical Data Sets

To acknowledge, as shown in Supplementary Table 2, a large majority of features were missing amongst sites, thus findings should be viewed as a large single institution retrospective study with external validation needed in the future. With the adoption of the EHR, patient-related data has exponentially grown but remains an under-used resource due to the complexity of mining data and uncovering novel associations. Such data provides the potential for generating new patient stratification strategies.<sup>30</sup> Our unsupervised PCA eigenvector approach may help rectify these challenges by providing easy-to-interpret visualizations of clinical data relationships. Illustrating this method's utility in our study, the vectors measuring radiation metrics showed opposite directionality to features measuring patient age. The directionality of these vectors thus underscores the known clinical management of GB patients whereby older individuals often receive hypofractionated radiation dosage due to toxicity.<sup>31</sup> The application of our eigenvector plot to initially detect relationships of patient survival to CBCs illustrates the utility of applying this method in large clinical data sets as a first-pass visualization approach in identifying novel relationships to explore in a clinical study.

### CBC Stratification can be Applied to Identify Patients with Poorer Predicted Survival Outcome

It was highlighted that CBC metrics—namely WBC and neutrophil count—were inversely related to OS and PFS. Interestingly, past studies have indicated that several components of CBC tests have predictive outcomes in OS. Namely, Pierscianek et al. and Jarmuzek et al. both retrospectively identified similar effects of WBC counts as prognostic factors in OS using CBCs collected during admission or pre-operatively for a potential glioma.<sup>32,33</sup> Nevertheless, these studies do not recapitulate the relevant timepoint within our study which suggested WBC counts collected prior to initiating ChemoRT as the most predictive of survival. To this point, Schernberg et al. similarly assessed the utilization of CBCs during pretreatment for ChemoRT and found neutrophilia, advanced age, and more complete resection as features that independently decreased OS in a multivariate model, while steroid consumption did not.<sup>34</sup> These findings largely parallel our Cox regression models which found patient age, ChemoRT treatment regime, and WBC count as predictive of OS.

It was noted as well that the median survival time in the WBC-Hi group was approximately 12 months while the WBC-Lo group was 18 months. Although the coverage of patients in our study ranges over the past decade, studies have reported the median survival of GB ranging from 12 to 15 months with ChemoRT treatment.<sup>35–37</sup> In turn, it may be suggestive that patients with lower WBC load experience better survival outcomes than just those with neutrophilia declining more rapidly. Conversely, assessing the cutoffs used for our WBC (Lo < 6.70; Hi > 11.90) and neutrophil (Lo < 4.40; Hi > 9.26) groupings, our low cutoffs are contained in the normal CBC reference range, but high cutoffs exceed the upper limit of normal,<sup>38</sup> Thus patients with elevated WBC at ChemoRT planning may warrant more careful monitoring.

The relationship between survival and WBC count additionally raises applications to future therapeutic treatment of GBs—namely immunotherapies. Although effective in other solid tumors, the success of PD-1/PD-L1 immune checkpoint inhibitors and dendritic cell vaccines have been marginal in GB.<sup>39–41</sup> Contrary to beliefs that GBs were largely immune-privileged tumors, growing evidence supports robust recruitment of pro-tumor-immune cell populations such as tumor-associated neutrophils and polymorphonuclear myeloid-derived suppressor cells.<sup>42,43</sup> Thus, our poorly survived patients with elevated WBC/neutrophil load may underscore these biological mechanisms of tumor progression. Although future exploration is needed, such routine markers may be critical in identifying patients who are poor candidates for immunotherapy treatment due to unfavorable tumor microenvironment.

### Fast Tapering of Steroids following Surgery May Influence Long Term Outcomes of Patients

An important consideration that should be highlighted in our study was the evidence that prior to ChemoRT treatment, a large variance in steroid tapering and functional status (KPS) was noted across our study groups. Namely, while studies have evidenced that WBC prediction is independent of corticosteroid use statistically, the biological influence of corticosteroids on both neutrophil and lymphocyte count has been long recognized.<sup>44,45</sup> Alternative studies, like Dubinski et al., have directly implicated that the administration of dexamethasone induces leukocytosis which was associated with poor survival.<sup>46</sup> Conversely, however, as our study found poorer KPS in the WBC-Hi group, these differences in steroid dosing may reflect more advanced disease. Nevertheless, although steroids have been long used to provide supportive therapy, these findings may suggest the need to set a more consistent standard of quickly tapering patients off steroids or identifying who benefits from a prolonged course.<sup>47</sup> Our observed WBC counts and PD-L1 measure may be suggestive of which patients in fact need to have steroid doses modified due to influence on immune cell activity. Studies have evidenced that the increased administration of steroids promotes immune cell dysfunction, namely in T-cell compartments, by promoting increased expression of PD-L1 in the microenvironment that advances

dysregulation of the immune response.<sup>48,49</sup> Although these findings may complicate the utilization of CBC measures for survival prediction, these observations, more importantly, underscore the need to better consider the use of steroids for the symptomatic relief of GB patients. Although steroids may be used for patients with worse functional status, steroids may conversely worsen symptoms by causing immune cell dysfunction.

## Conclusions

The use of data derived from the EHR will remain a powerful resource. However, the sheer complexity of both collecting and assaying data to uncover novel research discoveries poses a challenge for future studies. The deliverables of our study include our standardized REDCap data collection form for evaluation of GB, an unsupervised analysis framework to initially explore clinical data sets through PCA eigenvector visualization, and an automated image analysis pipeline for PD-L1 staining. Furthermore, the measure of CBC load at pretreatment for ChemoRT can be applied to identify patients at risk for unfavorable survival due to high WBC load coupled with elevated PD-L1 staining. However, deeper exploration and validation of these relationships and the tapering of steroids are important considerations in future studies of GB patient management.

## Supplementary Material

Supplementary material is available online at *Neuro-Oncology* (<https://academic.oup.com/neuro-oncology>).

## Keywords

clinical decision-making | glioblastoma | machine learning | survival outcomes

## Funding

J.J.O. acknowledges the National Institute of Neurological Disorders and Stroke (NINDS) R03NS116334. J.J.O. and W.W. acknowledge support from the Ohio State University-Translational Data Analytics Institute (OSU-TDAI) seed grant and the National Center for Advancing Translational Sciences (NCATS) TL1TR002735 and UL1TR002733. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflict of interest statement

None declared.

## Author Contributions

W.W., Z.T.K., C.H., I.Z., G.T., H.M., J.K., M.G.M., and K.E.B. were involved in the collection and analysis of data presented in the study. W.W., G.T., R.M.R., H.M., M.D.A., M.O.C., S.B., J.B.L., P.G., and J.J.O. were involved in the design and interpretation of the work. All authors read and approved the final manuscript.

## Data Availability

The data sets used and/or analyzed during the current study are available from the corresponding authors upon reasonable request. Dissemination of the applied REDCap form can be found in supplement or supplied from corresponding authors.

## References

- Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. *Neuro-Oncology*. 2021;23(Supplement\_3):iii1–iii105.
- Silbergeld DL, Rostomily RC, Alvord EC, Jr. The cause of death in patients with glioblastoma is multifactorial: clinical factors and autopsy findings in 117 cases of supratentorial glioblastoma in adults. *J Neurooncol*. 1991;10(2):179–185.
- Stupp R, Mason WP, van den Bent MJ, et al.; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987–996.
- Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology*. 2021;23(8):1231–1251.
- Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med*. 2005;352(10):997–1003.
- Gomes I, Moreno DA, Dos Reis MB, et al. Low MGMT digital expression is associated with a better outcome of IDH1 wildtype glioblastomas treated with temozolomide. *J Neurooncol*. 2021;151(2):135–144.
- Ma S, Rudra S, Campian JL, et al. Prognostic impact of CDKN2A/B deletion, TERT mutation, and EGFR amplification on histological and molecular IDH-wildtype glioblastoma. *Neuro-Oncology Advances*. 2020;2(1):vdaa126.
- Heimberger AB, Hlatky R, Suki D, et al. Prognostic effect of epidermal growth factor receptor and EGFRvIII in glioblastoma multiforme patients. *Clinical Cancer Res Official J Am Assoc Cancer Res*. 2005;11(4):1462–1466.
- Cevik L, Landrove MV, Aslan MT, et al. Information theory approaches to improve glioma diagnostic workflows in surgical neuropathology. *Brain Pathol*. 2022;32(5):e13050.
- Wang W, Howard D, Giglio P, Thomas D, Javier Otero J. Bioethical implications of current state practices of molecular diagnostics in neuropathology. *Neuro-Oncology*. 2022;24(6):853–854.

11. Dundas NE, Ziadie MS, Revell PA, et al. A lean laboratory: operational simplicity and cost effectiveness of the Luminex xTAG™ respiratory viral panel. *J Mol Diagn*. 2011;13(2):175–179.
12. Gorlia T, van den Bent MJ, Hegi ME, et al. Nomograms for predicting survival of patients with newly diagnosed glioblastoma: prognostic factor analysis of EORTC and NCIC trial 26981-22981/CE3. *Lancet Oncol*. 2008;9(1):29–38.
13. Johnson DR, Sawyer AM, Meyers CA, O'Neill BP, Wefel JS. Early measures of cognitive function predict survival in patients with newly diagnosed glioblastoma. *Neuro-Oncology*. 2012;14(6):808–816.
14. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381.
15. Harris PA, Taylor R, Minor BL, et al; REDCap Consortium. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383.
17. Tierney N, Cook D, McBain M, et al. Data Structures, Summaries, and Visualisations for Missing Data. R-Packages. Published online 2019. <https://CRAN.R-project.org/package=naniar>
18. Buuren S van, Groothuis-Oudshoorn K, Vink G, et al. *mice: Multivariate Imputation by Chained Equations*. R-Packages. Published online November 19, 2022. <https://CRAN.R-project.org/package=mice>
19. Kassambara A, Mundt F. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R-Packages. Published online April 1, 2020. <https://CRAN.R-project.org/package=factoextra>
20. Wang W, Alzate-Correa D, Alves MJ, et al. Machine learning-based data analytic approaches for evaluating post-natal mouse respiratory physiological evolution. *Res Physiol Neurobiol*. 2021;283:103558.
21. Therneau TM, Lumley TL, Elizabeth A, Cynthia C. *survival: Survival Analysis*. R-Packages. Published online March 3, 2022. <https://CRAN.R-project.org/package=survival>
22. Kassambara A, Kosinski M, Biecek P, Fabian S. *Drawing Survival Curves using “ggplot2”* [R package survminer version 0.4.5]. R-Packages. Published online 2019. <https://CRAN.R-project.org/package=survminer>
23. Hadley Wickham. *Ggplot2 Elegant Graphics for Data Analysis*. Cham Springer International Publishing; 2016.
24. Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*. 2010;26(7):979–981.
25. van der Walt S, Schönberger JL, Nunez-Iglesias J, et al.; scikit-image contributors. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453.
26. Kalinowski T, Ushey K, Allaire JJ, et al. *reticulate: Interface to “Python.”* R-Packages. Published online January 27, 2023. <https://CRAN.R-project.org/package=reticulate>
27. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62–66.
28. Tin KH. Random decision forests. *IEEE Xplore* 1995;1(1):278–282.
29. Igarashi T, Teramoto K, Ishida M, Hanaoka J, Daigo Y. Scoring of PD-L1 expression intensity on pulmonary adenocarcinomas and the correlations with clinicopathological factors. *ESMO Open*. 2016;1(4):e000083.
30. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395–405.
31. Perry JR, Laperriere N, O'Callaghan CJ, et al.; Trial Investigators. Short-course radiation plus temozolomide in elderly patients with glioblastoma. *N Engl J Med*. 2017;376(11):1027–1037.
32. Pierscianek D, Ahmadipour Y, Michel A, et al. Preoperative survival prediction in patients with glioblastoma by routine inflammatory laboratory parameters. *Anticancer Res*. 2020;40(2):1161–1166.
33. Jarmuzek P, Kot M, Defort P, et al. Prognostic values of combined ratios of white blood cells in glioblastoma: a retrospective study. *J Clin Med*. 2022;11(12):3397.
34. Schernberg A, Nivet A, Dhermain F, et al. Neutrophilia as a biomarker for overall survival in newly diagnosed high-grade glioma patients undergoing chemoradiation. *Clinic Translation Rad Oncol*. 2018;10:47–52.
35. Brown NF, Ottaviani D, Tazare J, et al. Survival outcomes and prognostic factors in glioblastoma. *Cancers*. 2022;14(13):3161.
36. Johnson DR, O'Neill BP. Glioblastoma survival in the United States before and during the temozolomide era. *J Neuro-Oncol*. 2011;107(2):359–364.
37. Marengo-Hillebrand L, Wijesekera O, Suarez-Meade P, et al. Trends in glioblastoma: outcomes over time and type of intervention: a systematic evidence based analysis. *J Neuro-Oncol*. 2020;147(2):297–307.
38. Tefferi A, Hanson CA, Inwards DJ. How to interpret and pursue an abnormal complete blood cell count in adults. *Mayo Clin Proc*. 2005;80(7):923–936.
39. Reardon DA, Omuro A, Brandes AA, et al. OS103 randomized phase 3 study evaluating the efficacy and safety of nivolumab vs bevacizumab in patients with recurrent glioblastoma: CheckMate 143. *Neuro-Oncology*. 2017;19(Suppl 3):iii21–iii21.
40. Lim M, Weller M, Idbaih A, et al. Phase III trial of chemoradiotherapy with temozolomide plus nivolumab or placebo for newly diagnosed glioblastoma with methylated MGMT promoter. *Neuro-Oncology*. 2022;24(11):1935–1949.
41. Liao LM, Ashkan K, Tran DD, et al. First results on survival from a large Phase 3 clinical trial of an autologous dendritic cell vaccine in newly diagnosed glioblastoma. *J Transl Med*. 2018;16(1):142.
42. Khan S, Mittal S, McGee K, et al. Role of neutrophils and myeloid-derived suppressor cells in glioma progression and treatment resistance. *Int J Mol Sci*. 2020;21(6):1954.
43. Veglia F, Sanseviero E, Gabrilovich DI. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. *Nat Rev Immunol*. 2021;21(8):485–498.
44. Boggs DR, Athens JW, Cartwright GE, Wintrobe MM. The effect of adrenal glucocorticosteroids upon the cellular composition of inflammatory exudates. *Am J Pathol*. 1964;44(5):763–773.
45. Mishler JM, Emerson PM. Development of neutrophilia by serially increasing doses of dexamethasone. *Br J Haematol*. 1977;36(2):249–257.
46. Dubinski D, Won SY, Gessler F, et al. Dexamethasone-induced leukocytosis is associated with poor survival in newly diagnosed glioblastoma. *J Neurooncol*. 2018;137(3):503–510.
47. Roth P, Happold C, Weller M. Corticosteroid use in neuro-oncology: an update. *Neuro-Oncol Pract*. 2014;2(1):6–12.
48. Herold MJ, McPherson KG, Reichardt HM. Glucocorticoids in T cell apoptosis and function. *Cell Mol Life Sci*. 2005;63(1):60–72.
49. Dietrich J, Rao K, Pastorino S, Kesari S. Corticosteroids in brain cancer patients: benefits and pitfalls. *Expert Rev Clin Pharmacol* 2011;4(2):233–242.