Methodology matters: characterization of glioma through advanced MR imaging

Schmitz Abecassis, B.

7 MRI phenotypes of glioblastomas early after treatment are suggestive of overall patient survival

Bárbara Schmitz-Abecassis
Linda Dirven
Janey Jiang
Jasmin A. Keller
Robert J. I. Croese
Daniëlle van Dorth
Rashid Ghaznawi
Ilse M. J. Kant
Martin J. B. Taphoorn
Matthias J. P. van Osch
Johan A. F. Koekkoek
Jeroen de Bresser

Neuro-Oncology Advances (2023). DOI:10.1093/noajnl/vdad133

7.1 Abstract

Distinguishing true tumor progression (TP) from treatment induced abnormalities (e.g. pseudo-progression (PP) after radiotherapy) on conventional MRI scans remains challenging in patients with a glioblastoma. We aimed to establish brain MRI phenotypes of glioblastomas early after treatment by combined analysis of structural and perfusion tumor characteristics, and assessed the relation with recurrence rate and overall survival time.

Structural and perfusion MR images of 67 patients at 3 months post-radiotherapy were visually scored by a neuroradiologist. In total 23 parameters were predefined and used for hierarchical clustering analysis. Progression status was assessed based on the clinical course of each patient 9 months after radiotherapy (or latest available). Multivariable Cox regression models were used to determine the association between the phenotypes, recurrence rate and overall survival.

We established 4 subgroups with significantly different tumor MRI characteristics, representing distinct MRI phenotypes of glioblastomas: TP and PP rates did not differ significantly between subgroups. Regression analysis showed that patients in subgroup 1 (characterized by having mostly small and ellipsoid nodular enhancing lesions with some hyper-perfusion) had a significant association with increased mortality at 9 months (HR:2.6 (CI:1.1–6.3); p=0.03) with a median survival time of 13 months (compared to 22 months of subgroup 2).

Our study suggests that distinct MRI phenotypes of glioblastomas at 3 months post-radiotherapy can be indicative of overall survival, but does not aid in differentiating TP from PP. The early prognostic information our method provides might in the future be informative for prognostication of glioblastoma patients.

7.2 Importance of the study

Determining brain MRI phenotypes of glioblastomas early after treatment can help in showing which combination of MRI markers is driving a lower survival chance 9 months after treatment. These distinct MRI phenotypes of glioblastomas could in the future guide complex clinical decision making based on patient prognosis early after treatment assessment.

7.3 Key points

Distinct MRI phenotypes of glioblastomas at 3 months post-radiotherapy are significantly associated with overall survival.

The same MRI phenotypes at 3 months post-radiotherapy do not seem to aid in differentiating between true- and pseudo-progression at 9 months after radiotherapy.

7.4 Introduction

Glioblastoma is the most common and severe type of primary malignant brain tumor[165]. Current multimodal treatment after surgical resection includes radiotherapy and concomitant and adjuvant chemotherapy with temozolomide[156]. Despite this treatment, a high local recurrence rate is observed during the disease course (90%)[166]. MRI is the cornerstone for brain tumor surveillance, and aids clinicians in guiding management decisions. However, a challenge is that high dose radiotherapy may cause treatment induced abnormalities in the early stages after treatment (i.e. pseudo-progression (PP)), which can look similar to tumor progression on conventional MRI scans[156,167]. Therefore, distinguishing between true tumor progression (TP) and PP early after treatment can be challenging. Early differentiation could aid clinicians in accurately identifying patients who require an alternative treatment strategy to delay further disease progression, and at the same time spare responding patients the burden of additional tumor-targeted treatment.

Structural MRI markers have shown some added value for identifying TP, specifically when assessing the size and the morphology of the enhancing lesion on post contrast $3DT_1$ images[168–171]. Furthermore, several studies have shown that perfusion MRI (with dynamic susceptibility contrast (DSC) and/or arterial spin labeling (ASL)) have added value in distinguishing between TP and PP in glioblastomas[30,172–177]. However, despite perfusion MRI techniques being promising in differentiating between TP and PP, individual MRI markers showed at best only a modest association with tumor progression and overall survival. This indicates the need to combine MRI markers to get a more reliable early assessment of TP in glioblastomas[177].

More recently, a number of studies have focused on the use of radiomics in glioblastomas to analyze MRI markers in a combined way. One previous study found that structural MRI markers

(gray level texture markers) were associated with TP[178]. Few previous studies have specifically applied radiomics models on perfusion MRI with the aim of developing models that could accurately predict TP[179–181]. Moreover, radiomics approaches such as deep learning models rely on MRI markers of higher order that are not directly clinically translatable, and it is not always completely clear how these markers drive the algorithms decision making (i.e., which combination of MRI markers). With this in mind we set out to explore an alternative approach, which includes grouping of patients with glioblastoma based on clinically scored structural and both ASL and DSC perfusion MRI markers[182,183]. We subsequently studied how these sub-groups progressed over time. We have previously developed a method that was able to identify brain MRI phenotypes based on hierarchical clustering, which resulted in clinical meaningful sub-categories in other disease conditions. For example, we identified brain MRI phenotypes related to predisposition to post operative delirium (in preoperative patients) and brain MRI phenotypes related to an increased risk of stroke and mortality (in patients with manifest arterial disease)[184]. To date, it is unknown what specific MRI phenotypes exist in patients with a glioblastoma.

We hypothesize that MRI phenotypes of glioblastomas based on both structural and perfusion tumor characteristics early after treatment could help in the risk assessment of glioblastoma recurrence rate and overall survival time. We therefore aimed to establish brain MRI phenotypes of glioblastomas early after treatment by combined analysis of radiological scores of structural and perfusion tumor characteristics, and to assess the relation of these phenotypes with tumor recurrence rate and overall survival time.

7.5 Materials and methods

7.5.1 Patient population

Patient clinical data were retrospectively retrieved from the clinical archive of the Leiden University Medical Center between the period of January 2015 and February 2022 following local IRB regulations. The study population consisted of adult patients with a histologically confirmed grade IV glioblastoma IDH wild-type or diffuse astrocytoma IDH-mutant, following the most recent WHO guidelines at the time of diagnosis. Consecutive patients who received postoperative treatment consisting of radiotherapy (in combination with concomitant and adjuvant chemotherapy), with at least a 3 months post-radiotherapy follow-up MRI scan with both ASL and DSC scans, and confirmation of TP or PP were included.

7.5.2 Tumor progression and survival assessment

The diagnosis of TP or PP was based on the patients' medical charts including clinical and radiological findings discussed in the multidisciplinary team meetings at either 3, 6 and/or 9 months after radiotherapy. The diagnosis at each of the time points was scored on a 5-point Likert scale: 1) Definite PP; 2) probable PP; 3) no preference ; 4) probable TP; 5) definite TP. The conversion to a binary scale was done by assessing scores 1 and 2 as tumor progression and 4 and 5 as no progression. In this way, the binary score agrees with clinical practice, i.e., if there was doubt about the progression status (score of 3), treatment was continued (and assumed there was no progression). Updated molecular and pathological findings were leading regarding the diagnosis in case of a re-resection. At the latest timepoint available (maximum of 9 months after radiotherapy), patients were only considered to have tumor progression if this was suggested by the clinical and radiological assessment, if the anti-tumor treatment regimen was changed or if the patient was deceased.

Patient survival was calculated as the time between the start of tumor-targeted therapy (i.e. the day of tumor surgical resection or biopsy) and the date of death.

7.5.3 Type of surgical resection

All patients underwent surgery for either maximally safe tumor removal (i.e., total or partial resection of the enhancing part of the tumor) or a biopsy to obtain a histopathological diagnosis. Patients were considered to have had a total resection of the enhancing parts if no enhancing lesion was observed on the directly postoperative MRI scan (performed within 48 hours after tumor resection). In contrast, if residual tumor enhancement was found on the directly postoperative MRI scan, it was considered a partial resection.

7.5.4 MRI scans

Patients were scanned at approximately 3 months post-radiotherapy on a 3T MR scanner (Philips Ingenia or Achieva, Philips Healthcare, Best, The Netherlands). The MRI scans acquired followed the routine clinical guidelines for all scans collected, thus all imaging acquisition parameters reported are based on standard clinical practice. These included a post gadolinium contrast enhanced $3DT_1$ scan with 3D-TFE readout and T_2-FLAIR scan, acquired with the following parameters: $3DT_1$: TR = 9.91ms, TE = 4.67ms, resolution = 1x1x1mm, field of view (FOV) = 220x175x156mm, 0.3 mL per kg bodyweight of gadolinium-based contrast agent (gadoterate meglumine) and T_2-FLAIR: TR = 11000 ms, TE = 125ms, resolution = 0.4x0.4x5.5mm, FOV = 220x175mm. During the period of data collection, changes to the ASL protocol were made, including transitioning from a 2D to 3D pCASL having changed the label duration (LD) and post-label duration (PLD). For 2D pCASL, LD = 1650ms and PLD = 1525ms (first slice) & 2120ms (last slice). Whereas for the 3D pCASL both LD and PLD were 1800ms. The remainder of parameters included resolution (2D/3D) = 3x3mm/4x4mm, slice thickness (2D/3D) = 7mm/6mm, FOV = 240x240mm. Finally the DSC scans were acquired with a SE-EPI sequence with the following parameters, TE = 75ms, TR = 1600ms, resolution = 2.6x2.3x5mm, FOV = 240x210mm; a third of the contrast agent was injected as pre-bolus.

7.5.5 Radiological scoring of the brain MRI scans

An independent neuroradiologist scored the structural brain scans, i.e. the contrast enhanced 3D T_1 and the T_2-FLAIR scans of the 3 months post-radiotherapy visit (and in doubt consulted a second experienced neuroradiologist for consensus), to determine 1) whether the tumor contrast enhancing lesions were either nodular or patchy on the contrast enhanced 3D T_1, 2) the presence of T_2 hyperintense areas surrounding the enhancing lesion on the T_2-FLAIR and 3) the size of the tumor contrast enhancing and T_2 hyperintense area in 3 orthogonal directions. The T_2 hyperintense area was defined as the confluent hyperintense signal surrounding the contrast enhancing lesion, excluding any resection cavities. The measurements in 3 orthogonal directions were used to estimate the tumor volume for the contrast enhancing lesions and T_2 hyperintense areas individually, as well as to calculate the shape as the eccentricity factor (EF). The volume was calculated using an ellipse formula, as this has been shown to correlate well with the absolute tumor volume, using the following formula[185]:

$$\frac{4}{3}\pi\ D1\ D2\ D3$$

Where D1, D2 and D3 correspond to the sagittal, coronal and axial measurements, respectively. The resulting volume was afterwards converted to milliliters for the final volume calculation. The EF was calculated according to the following formula:

$$EF = \sqrt{1 - \left(\frac{PPD}{MD}\right)^2}$$

Where MD is the maximal diameter (highest value in all 3 directions) and PPD is the largest perpendicular diameter, i.e. the largest diameter in the other two directions[186].

Tumor lesion perfusion was scored qualitatively using the contrast enhanced $3DT_1$ and the T_2-FLAIR as anatomical references. On the DSC relative cerebral blood volume (rCBV) maps were scored as either increased (hyper-perfusion), decreased (hypoperfusion) or no change (iso-perfusion) compared to contralateral normal tissue. The ASL was scored as hyperperfusion or isoperfusion compared to the contralateral side, since hypoperfusion on ASL scans is especially difficult to identify[187]. Perfusion scores for both contrast enhancing lesions and T_2 hyperintense areas were separately determined. For the clustering analysis, if there was more than one lesion in a patient, the most aggressive tumor contrast enhancing lesion and related T_2 hyperintense area were included per patient. The lesion with the largest volume, increased perfusion and most nodularity was considered to be the most aggressive. In total, 23 radiological tumor markers were included (see the supplementary methods). These markers were rigorously selected to be the most representative to avoid overfitting and selection bias of the model.

7.5.6 Statistical analysis

7.5.6.1 Hierarchical clustering analysis

To identify different brain MRI phenotypes at 3 months post-radiotherapy, hierarchical clustering was performed on the available patient data consisting of the radiological structural and perfusion tumor characteristics. The visually scored markers were included in the model as either binary, categorical or continuous variables. Continuous markers that did not have a normal distribution, including the volume and eccentricity, were normalized by multiplying by 100 and natural log-transforming, and thereafter normalized into z-scores. The normalization step allowed all markers to be equally scaled and then weighted by the analysis model.

Hierarchical clustering was performed using the Ward's method, Nbclust[188], factoextra[189], cluster[190] and dendextend[191] in R version 4.1.2 (R Core-Team 2021). Initially the model considers each patient as an individual cluster after which it tries to iteratively merge the two closest clusters while equally weighing each marker (i.e., merge the clusters that share the highest number of markers in common; the agglomerative approach). In this way the clustering analysis is performed in a hierarchical manner. Each time sub-clusters are merged into one cluster, the distance between the remaining clusters is subsequently updated in the model, and the next iteration starts. Ultimately, when all clusters have been merged only one cluster will remain. To determine the optimal number of clusters from the hierarchical clustering analysis, we used the heatmap and Dunn index. The Dunn index calculates the ratio of the smallest distance between markers within each cluster over the maximum distance between the clusters.

7.5.6.2 Differences between the subgroups with a different MRI phenotype of glioblastomas

Descriptive statistics were used to describe the patient population. Clinical characteristics between the subgroups identified with the hierarchical cluster analysis were compared as

follows: for categorical and binary variables, a chi-square test was used; continuous variables were compared using a one-way ANOVA or Kruskal Wallis test, depending on the distribution of the tested variable. In order to assess how the subgroups differed from each other considering the different variables, Bonferroni post-hoc analyses were performed for continuous variables and Bonferroni chi-square residual analyses for categorical variables.

7.5.6.3 Differences between the overall survival time of the subgroups

A log-rank test was performed to determine whether the median overall survival times were significantly different between the different subgroups. For this we compared the median overall survival times between subgroups 1, 3 and 4 and compared to the reference one, subgroup 2.

7.5.6.4 Association between different MRI phenotypes of glioblastomas and progression and survival

First, a multivariable logistic regression analysis was performed to assess the association between MRI phenotype subgroups and TP. The model was corrected for age, KPS score and surgical resection type. Whereas the first two variables were continuous, surgical resection type was categorized into three different groups (ie. biopsy, partial resection and total resection).

Secondly, an adjusted Cox proportional hazard model was used to determine the association between the MRI phenotype subgroups and overall survival time. The model was corrected for age, KPS score and surgical resection type. For sensitivity analysis, an unadjusted multivariable Cox proportional hazard model was performed to assess the influence of correcting for clinical variables on our results and thus to get more insight into the data. The subgroup chosen as reference for the regression analysis had the least aggressive MRI markers (i.e., subgroup 2). This subgroup with the least aggressive radiological markers was identified by an experienced neuroradiologist who assessed the structural and perfusion radiological markers of all subgroups (e.g. hyper-perfusion and increased nodular enhancement were considered aggressive markers).

The threshold for significance was $p \leq 0.05$. All statistical analyses were performed using IBM SPSS version 25 (Chicago-IL).

7.5.7 Data availability

Data will be made available upon reasonable request.

7.6 Results

7.6.1 Patient population

In total, 67 patients with a glioblastoma were eligible according to our inclusion criteria and were included in the analysis (Table 1). The majority (72%) had an Isocitrate Dehydrogenase (IDH)-wildtype tumor and most (61%) patients were male, with a mean age of 60 years (standard deviation 13 years). Most patients (79%) had undergone total or partial resection, radiotherapy as well as concomitant and adjuvant chemotherapy, and a small group of patients were only treated with radiotherapy (21%). See Table 1 for all sociodemographic and clinical variables.

Representative examples of the scored MRI markers can be found in Figure 1, which includes perfusion markers characterized in ASL and rCBV maps as well as structural markers such as enhancing lesion patterns on contrast enhanced $3DT_1$ and T_2 hyperintense areas on T_2-FLAIR.

7.6.2 Identification of subgroups with a different MRI phenotype of glioblastomas

The hierarchical clustering analysis resulted in the heatmap displayed in Figure 2. Establishing the optimal number of subgroups was a multistep process. First, we considered that with our number of patients included, the number of subgroups should be relatively low. Second, we inspected the heatmap (Figure 2) and found a good between-subgroup separation with four subgroups. Lastly, we considered the Dunn index (DI), which also showed a relatively high value at four subgroups. The number of subgroups of patients with a different MRI phenotype of glioblastomas was therefore determined at four (with n= 12, 13, 17 and 25 patients in the subgroups respectively).

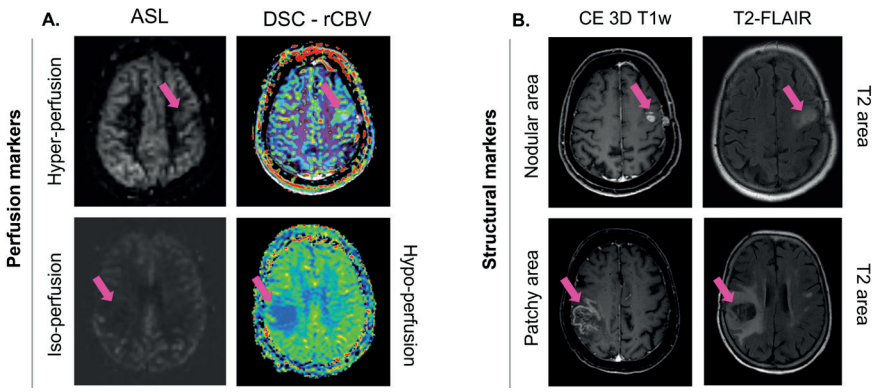


Figure 1. Examples of some of the MRI markers scored by a neuroradiologist.
A. Perfusion MRI markers retrieved from corresponding images, namely ASL and DSC-rCBV maps. On the upper and lower row examples of patients with increased and no change/decreased perfusion, for ASL and DSC respectively. In B. contrast enhanced (CE) $3DT_1$ and T_2-FLAIR MRI scans from which the structural markers were assessed; it shows examples of nodular and patchy contrast enhancing areas, as well as T_2 hyperintense areas. The scans on each row correspond to one patient.

Table 1. Baseline clinical characteristics of the patients with a glioblastoma.

Total number of patients included	**67**
Age	60 ± 13
Female	26 (39%)
Male	41 (61%)
IDH status	
Wild-type: glioblastoma	48 (72%)
Mutant: diffuse astrocytoma grade IV	5 (7%)
Unknown / not determined*	14 (21%)
MGMT status	
Methylated	18 (27%)
Unmethylated	49 (73%)
KPS median (range)	90 (60 - 100)

Surgery type	
Total Resection	31 (46%)
Partial Resection	22 (33%)
Biopsy	14 (21%)
Radiotherapy (total dose)	
40 Gy	10 (15%)
45 Gy	10 (15%)
60 Gy	47 (70%)
Temozolomide chemotherapy	**53 (79%)**

*Cases diagnosed prior to WHO 2016 classification. IDH: Isocitrate dehydrogenase; MGMT: O6-methylguanine-DNA methyltransferase; KPS: Karnofsky Performance Scale.

7.6.3 Differences between the subgroups with a different MRI phenotype of glioblastomas

A summary of the clinical characteristics of these patient subgroups can be found in Supplementary Material Table S1. There were no significant differences between subgroups regarding KPS score, radiotherapy dose received, gender, IDH status, MGMT promotor status and receiving combined chemoradiotherapy (all $p>0.05$). However, there was a significant difference in age and type of surgery between the 4 subgroups ($p=0.013$ and $p=0.001$ respectively). From the Bonferroni post-hoc analyses it was evident that the overall significant difference in age was particularly observed because patients in subgroup 3 were significantly older than in subgroup 1 (68 versus 54 years, respectively, $p=0.02$). For surgery type, subgroup 1 had significantly more total resections compared to subgroup 3 (75% versus 12%, respectively, $p=0.003$) and subgroup 4 (75% versus 12%, respectively, $p=0.001$), explaining the overall significant difference. Supplementary Table S2 shows the distribution of the MRI markers present in each subgroup. Regarding tumor location, the temporal tumor region was the only location that differed significantly between the subgroups, with more patients having a temporal lesion in subgroup 1 (50% versus 15%, 0% and 32% in groups 2, 3 and 4, $p=0.02$). As expected, the majority of all other MRI markers were significantly different between the resulting subgroups (all $p<0.05$) (Supplementary Table S2). The structural MRI markers that differed significantly between subgroups include the presence of a patchy enhancing lesions, the number of nodular and enhancing lesions, as well as the volume and eccentricity from both enhancing and T_2 hyperintense tumor lesions ($p<0.05$) (Supplementary Material Table S2). Moreover, the perfusion markers differed significantly between subgroups, such as the

DSC-rCBV for both nodular and patchy enhancing lesions, and the ASL-CBF of the nodular and patchy enhancing lesions and of the T_2 hyperintense tumor areas ($p<0.05$) (Supplementary Material Table S2).

Overall, subgroup 1 was characterized by relatively few, small and mostly nodular enhancing lesions with a more ellipsoid shape and some lesions showing hyper-perfusion. Subgroup 2 was characterized by relatively few, small lesions with mostly patchy enhancing lesions, with a more ellipsoid shape and almost no lesions with hyper-perfusion. Subgroup 3 had the most lesions that also had the highest volume and highest amount of lesions with hyper-perfusion. Lastly, subgroup 4 was characterized by a relatively moderate amount of nodular and patchy enhancing lesions with a relatively high volume, ellipsoid shape and moderate amount of lesions with hyper-perfusion. A summary of the most relevant MRI markers of the MRI phenotypes of glioblastomas can be found in Figure 4.

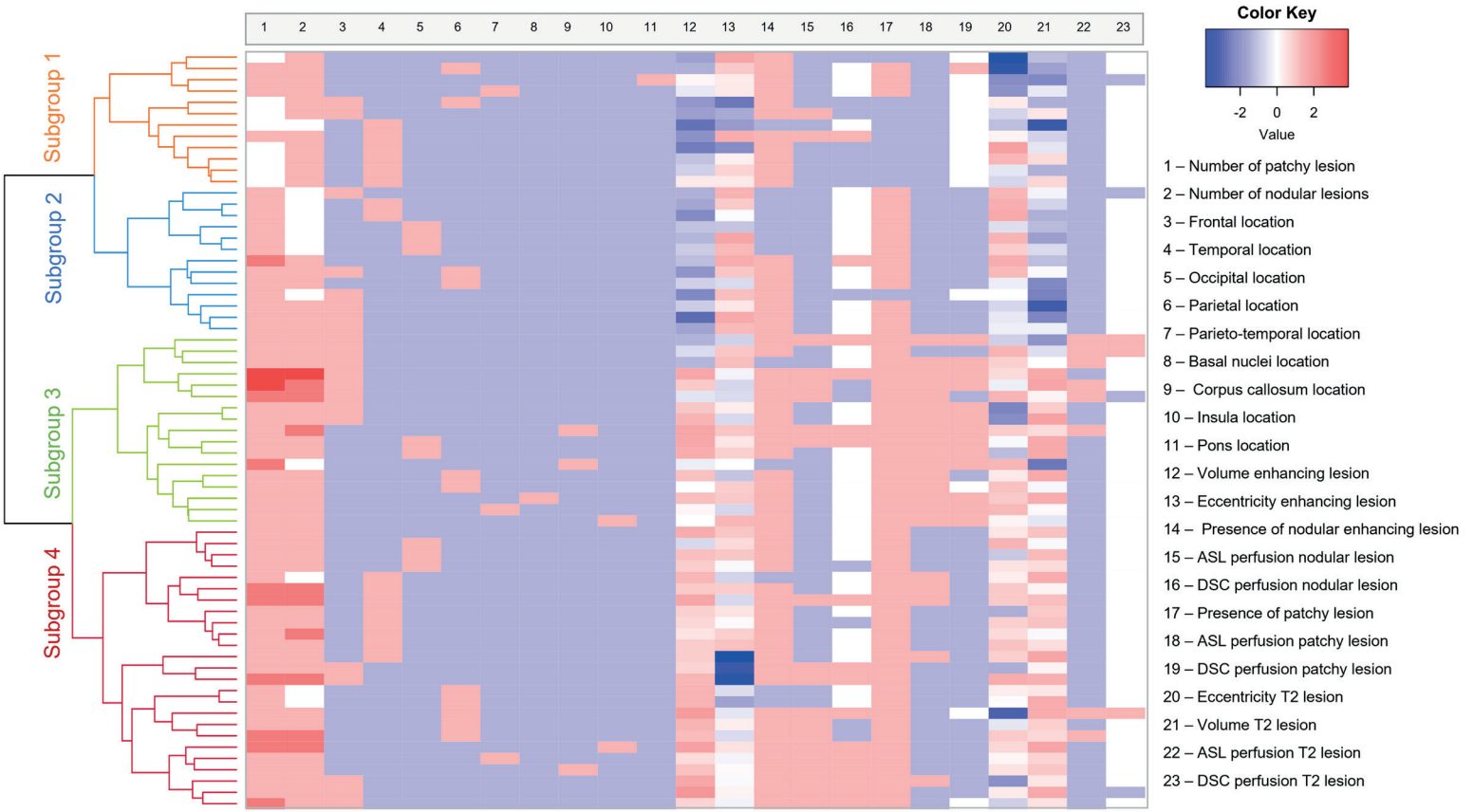


Figure 2. Heatmap results of the hierarchical clustering analysis.
The subgroups (subgroup 1 (n=12), subgroup 2 (n=13), subgroup 3 (n=17) and subgroup 4 (n=25)) are represented in different colors on the left side of the figure. Each row represents one patient and each column represents one MRI marker. In total 23 MRI markers were included. The normalized values were included, where blue represents a low value and red a high value of the MRI marker.

7.6.4 Differences between the overall survival time of the subgroups

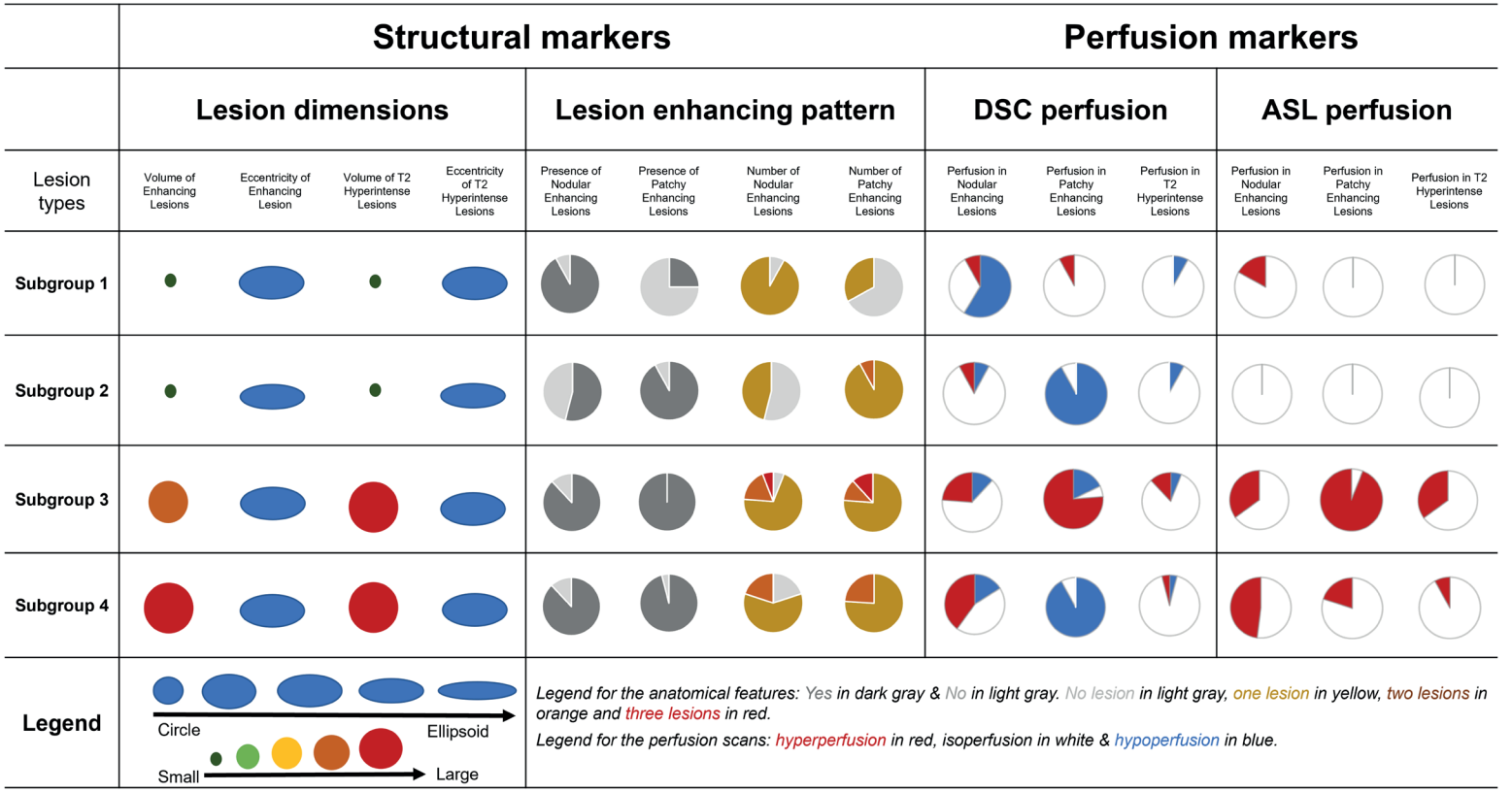
A significant difference was found between the overall survival times of the reference subgroup (subgroup 2) and subgroup 3 (p=0.009), but not between the reference subgroup and subgroup 1 (p=0.166) and 4 (p=0.191) (Supplementary Table S5, Supplementary Figure S3).

7.6.5 Association between different MRI phenotypes of glioblastomas and progression and survival

Out of the 67 patients, 49 showed a final diagnosis of TP at 9 months follow-up. Per subgroup this translated in 8/12 (67%) patients having TP in subgroup 1, 9/13 (69%) in subgroup 2, 14/17 (82%) in subgroup 3 and 18/25 (72%) in subgroup 4. From the multivariable logistic regression analysis, we found that subgroups 1 (HR: 1.3 (95% CI: 0.2 – 6.8); p = 0.772), 3 (HR: 1.7 (95% CI: 0.3 – 10.4); p = 0.570) and 4 (HR: 1.6 (95% CI: 0.3 – 8.4); p = 0.551) were not significantly associated with TP, compared to the reference subgroup 2 (Supplementary Table S6).

In total, 65 out of the 67 (97%) patients had passed away by the time the database was locked. The median survival in months per subgroup (with respective inter-quartile ranges) was: subgroup 1 = 13 (10 – 21) months, subgroup 2 = 22 (15 – 29) months, subgroup 3 = 11 (7 – 14) months, and subgroup 4 = 10 (8 – 18) months Figure 4 shows the results of the survival analyses where subgroup 2 was taken as the reference group. Our results show that subgroup 1 (HR: 2.6 (95% CI: 1.1 – 6.3); p=0.03) is significantly associated with mortality (Figure 5, Supplementary Figure S2), but not subgroups 3 and 4, when correcting for clinical variables (model 2). On the other hand, our secondary analysis included in the uncorrected logistic regression analysis, showed that subgroup 3 had a significant association with mortality (HR: 2.4 (95% CI: 1.1 – 5.0; p = 0.03), but not subgroups 1 and 4.

Figure 3. Illustration of the most representative structural and perfusion MRI markers as well as tumor volume and shape markers, per subgroup. These illustrative results are based on the values in Supplementary Table S2.



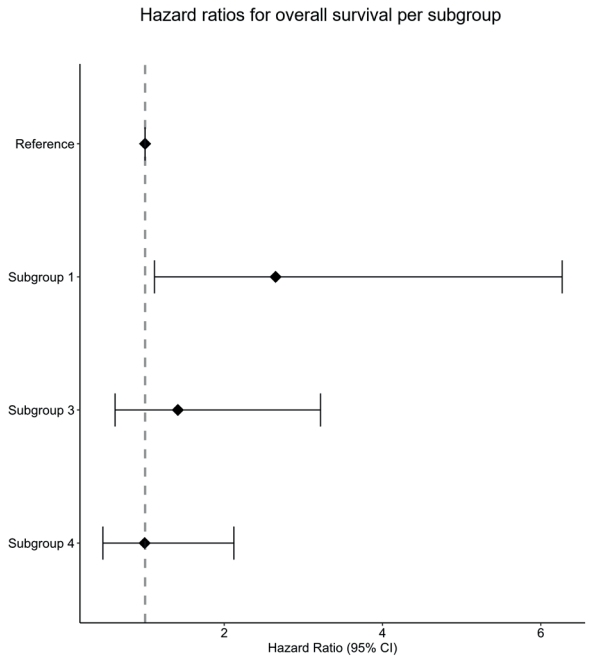


Figure 4. Forrest plot of the hazard ratios (95%-CI) per subgroup for survival. Illustrated results of the cox proportional hazards survival analysis are shown (adjusted for age, KPS and surgery type). Subgroup 2 has been set as a reference and marked with the striped line in red. Hazard ratios are shown with a 95% confidence interval.

7.7 Discussion

We identified distinct brain MRI phenotypes of glioblastomas in patients early after radiotherapy (at 3 months). Based on these brain MRI phenotypes we were able to define 4 distinct subgroups that also differed in their overall median survival times, but showed no differences in TP and PP rates.

In our current study, to establish different brain MRI phenotypes of glioblastomas, our clustering model hierarchically stratified the different patients based on the similarities between radiological MRI markers. Furthermore, we independently assessed the clinical outcomes. This approach, contrarily to most traditional machine learning models, is seen as a form of unsupervised learning. One of the advantages of our method includes the possibility of independently establishing new MRI marker combinations that in machine learning methods could otherwise not be determined.

The radiological MRI markers utilized in our study were based on radiological scoring of structural and perfusion images, and when combined, resulted in phenotypes of glioblastomas. Most of the MRI markers differed significantly between the 4 subgroups, and we identified subgroup 1 to have a significant association with mortality at 9 months. This subgroup's MRI phenotype of glioblastomas was characterized by relatively few, small and mostly nodular enhancing lesions with a more ellipsoid shape and some lesions showing hyper-perfusion, especially on ASL images, and the T_2 hyperintense area on the DSC scans showed mainly hypo-perfusion.

Some individual MRI markers that drive the model and that could influence the lower chance of overall survival in subgroup 1 can be identified. For example, the morphology of the nodular enhancing lesions are thought to reflect a more malignant phenotype[192]. Furthermore, the more ellipsoid shape of the lesions in subgroup 1 could indicate a more aggressive pattern of tumor infiltration in the neighboring healthy brain tissue[193]. Given the complexity of glioblastomas, and the uncertainty of what exact combination of MRI markers drives its severe prognosis, MRI phenotypes of glioblastomas can give more prognostically meaningful information and maybe in the future could help in treatment decisions.

Previous machine learning approaches have studied diagnostic performance using radiomics markers in brain gliomas[182,183]. For instance, a previous study utilized higher order texture and gray level intensity markers from ASL and DSC perfusion-weighted images and compared their quantitative patterns between high-grade and low-grade gliomas to assess diagnostic performance[182]. Radiomics in ASL and DSC was shown to be valuable, providing quantitative patterns to classify low and high grade gliomas with an area under the curve of 0.888 and 0.962, respectively[182]. Similarly, another previous study also investigated if, among others, grey level intensity and texture markers retrieved from DSC perfusion images, could identify glioma grades and IDH status. This study demonstrated stratification of glioma grades and IDH mutations status based on the DSC perfusion markers in a radiomics model performed correctly in 71% and 53% of the cases, respectively[183]. Both of these previous studies showed that a machine learning approach with MRI perfusion markers showed potential to reliably classify gliomas grades and molecular genotypes. In traditional machine learning approaches, markers of interest were pre-determined and coupled to a known outcome for predictive modeling, where the markers of interest are mostly abstract and of higher-order, not directly representing clinical MRI markers. However, the main disadvantage of these machine learning approaches is that it is not always clear which exact marker (or combination of markers) is associated with a certain clinical outcome. Contrarily, in the current study we did not train a model to predict a certain outcome. Rather, we are utilizing MRI markers that are more representative of clinical radiological markers, and group the patients according to how similar these markers are. After this grouping, post-hoc analyses is performed to gain insight in which combination of these markers (MRI phenotypes of glioblastomas) are underlying specific clinical outcomes. Utilizing phenotypes (i.e. a combination of markers) instead of single markers can be more advantageous, because it allows identification of which group of characteristics could be prognostically more meaningful in a disease with a complex biology. In this way our approach is unique and also allows for a more comprehensive link between phenotype and outcome, as was also shown in previous studies of our group in other diseases[184,194].

Regarding recurrence and overall survival outcomes of the 4 subgroups with a distinct MRI phenotype of glioblastomas, we observed different overall median survival times and TP cases. Subgroup 2 had the longest overall median survival time (22 months) and 69% of TP incidences. We found subgroup 1 to have a significant association with mortality at 9 months, while the percentage of TP cases (67%) was similar. These results were somewhat unexpected.

The discrepancy between TP cases and overall survival time at 9 months follow-up could be explained by the large overall survival time range, a possible effect of a few outliers cases, and the small sample size for each subgroup. Moreover, subgroup 1 was shown to have a significant association with mortality at 9 months follow-up when correcting for confounding variables, including age[195]. This suggests that such significant association observed in subgroup 1 is most likely a reflection of the tumor phenotype, not influenced by the age of the patients. When the model was not corrected for clinical variables, we found that only subgroup 3, with a considerably older median age (68 ± 11), had a significant association with mortality at 9 months. When comparing the median survival times, also not corrected for clinical confounders, a significant difference was also found for subgroup 3, as well as no significance difference for subgroups 1 and 4. The fact that clinical confounders are not weighted in most likely explains the non-significant findings, since both subgroups 1 and 4 have relatively short median survival times (compared to subgroup 2). An additional explanation for subgroup 4 could part from the crossing survival curves, which is probably due to a few outliers with longer survival times. Furthermore, we observed that patients with worse survival time in subgroups 3 and 4 (overall survival time of 11 and 10 months, respectively) not only had the lowest percentages of total resections (subgroup 3 and 4 = 12%), but also had the highest tumor volumes (subgroup 3 = enhancing lesion: 247.54 ml; T_2 lesion: 1321.70 ml and subgroup 4 = enhancing lesion: 374.45 ml; T_2 lesion: 1375.43 ml). This makes us belief that the subgroups not only reflect different tumor phenotypes, but also the differences in surgical treatment. It is known that resection type is an important prognostic factor for survival, but does not correlate with the occurrence of TP or PP (similar to what we have observed in the present study).

In clinical practice it is relevant to understand which demographic and disease specific characteristics play a role in prognosis of the patient's disease. Since glioblastomas are known to recur, early assessment of tumor characteristics could better indicate patient prognosis on a more personal level early after treatment. Contrary to traditional machine learning approaches, rather than helping predict an outcome, our study helps to understand which combination of radiological MRI markers are the imaging correlates of clinical outcomes, such as overall survival[196]. The early prognostic information our method provides might in the future be beneficial for prognosis of patients with a glioblastoma. More specifically, when assessing a patient's perfusion and structural tumor characteristics at 3 months, stratifying patients according to their MRI phenotype of glioblastomas could inform clinicians early on after treatment about the patient's outcome at 9 months. In order to further investigate the clinical impact, a larger study would need to be conducted, but our method shows promising results to justify and power such a study.

The strengths of our study include the well characterized radiological MRI markers of glioblastomas, including DSC as well as ASL perfusion markers in combination with the extensive follow-up availability of both clinical and MRI data[197]. This allowed us to determine the clinical outcome of all patients considering both survival and progression. Furthermore, since ASL is not widely implemented in clinical imaging guidelines of glioblastomas, this dataset is

unique for this patient population[16].

Our study also has some limitations. A first limitation is that our sample size was limited to 67 patients. A reason for this was that glioblastoma patients who were certain to have died from non-tumor related causes or did not have structural and perfusion MR data available because of MRI contraindications, were not included in our study. One of the evident consequences of our sample size is visible in the survival time ranges within subgroups. Our hierarchical clustering analysis approach could differentiate subgroups with primarily different radiological markers. These subgroups also showed to have different overall survival times. However, utilizing this approach for discriminating subgroups solely based on overall survival time is challenging. It could be interesting in the future to include a larger sample size and re-evaluate these findings. Although our selection procedure narrowed our patient population, it allowed us to be certain that the clinical outcome was related to their tumor diagnosis. Despite the relatively limited sample size we were able to find clinically meaningful associations. A second limitation of our study is that the IDH status for some patients was unknown, since some of the patients were diagnosed prior to 2016. This withheld us from including this variable in the survival and progression analyses, which would have been of added value since patients with an IDH mutation are known to have an overall better prognosis[198]. A third limitation is that over the years of data collection the parameters of the ASL MRI acquisition changed. However, these changes were only minimal (applied to only one patient), and were not expected to have affected the visual perfusion scoring by the neuroradiologist in a significant way. Since the data used was retrospectively collected, the ASL acquisition parameters were set according to clinical standards, only including one PLD. Although including more PLD could make the images less sensitive to ATT artifacts, this is not yet the standard in clinical practice. The goal of our study was to also investigate how perfusion markers in such a hierarchical clustering model could be used to establish glioblastoma phenotypes. We strived to utilize the maximum number of markers representative of clinical practice, without being redundant. However, a larger number of markers or a different selection of markers, which would also be representative of other biological processes in the tumor, could provide more complete phenotypes of this disease. This could also lead to overlapping markers and therefore pruning the model to these overlapping markers. It would be interesting to see the results of future studies which would, for instance, also include metabolic information derived from MR spectroscopy or chemical exchange saturation transfer. Lastly, the ratings of the MRI scans were performed by one experienced neuroradiologist who consulted a second experienced neuroradiologist when in doubt to obtain consensus. Using only limited raters could be perceived as a limitation. However, we chose to invest in the quality of the raters instead of the quantity of the raters to achieve high quality data.

In conclusion, we were able to establish 4 subgroups based on distinct brain MRI phenotypes of glioblastomas at 3 months post-radiotherapy. Our study suggests that these distinct MRI phenotypes of glioblastomas can be indicative of overall survival. The early prognostic information our method provides might be informative for prognosis in patients with a glioblastoma.

7.8 Acknowledgments

This research was funded by Medical Delta as part of the Cancer Diagnostic 3.0 program.

7.9 Supplementary information

7.9.1 Supplementary methods

Brain glioblastoma MRI markers were established from the visual scoring described in the methods section, resulting in the 23 markers described below. Markers 3 – 11, 14 and 17 were included in the model as binary variables; i.e. presence (1) or no presence (0) of the marker in the lesion being scored. Markers 1, 2, 15, 16, 18, 19, 22 and 23 were included as categorical variables; i.e. they were given a different number per classification (e.g. hypoperfusion -1, iso-perfusion 0 and hyper-perfusion 1). The remaining markers 12, 13, 20 and 21 were used as continuous variables. Note that one representative tumor enhancing area was chosen per patient in case of multiple ones. The tumor area with the most aggressive MRI markers was chosen (e.g. hyper-perfusion) and with presentation of a T_2 hyperintense area. The latter was the case to avoid having missing values in the model for variables considering T_2 hyperintense areas. The non-normally distributed continuous variables were transformed by multiplying by 100 and natural log-transformed to approximately conform to a normal distribution, and thereafter have their values normalized with z-scores.

1. Number of patchy areas
2. Number of nodular areas
3. Frontal location
4. Temporal location
5. Occipital location
6. Parietal location
7. Parieto-temporal location
8. Basal nuclei location
9. Corpus callosum location
10. Insula location
11. Pons location
12. Volume enhancing area
13. Eccentricity enhancing area
14. Presence of nodular enhancing area
15. ASL perfusion nodular area
16. DSC perfusion nodular area
17. Presence of patchy area
18. ASL perfusion patchy area

19. DSC perfusion patchy area
20. Eccentricity T_2 hyperintense area
21. Volume T_2 hyperintense area
22. ASL perfusion T_2 hyperintense area
23. DSC perfusion T_2 hyperintense area

7.9.2 Supplementary results

As a results of the clustering analysis, we obtained 4 subgroups. Each subgroup has a unique brain MRI phenotype of glioblastomas which showed specific structural and perfusion MRI characteristics. An extended description of the subgroups shown in Figure 4 can be found below:

Subgroup 1

With regards to the structural MRI markers, subgroup 1 has the least number of patchy enhancing (PE) areas, the smallest volume of enhancing and T_2 hyperintense areas and the most round shape of the enhancing and T_2 areas.

Considering DSC perfusion markers, subgroup 1 has the least amount of hyper-perfusion in nodular enhancing (NE) areas, shows mainly iso-perfusion in PE areas, and shows mostly hypo-perfusion in T_2 hyperintense areas.

Considering ASL perfusion markers, subgroup 1 has a relatively high amount of hyper-perfusion in NE areas, and only has areas with iso-perfusion in PE areas and in T_2 hyperintense areas.

Subgroup 2

With regards to the structural markers, subgroup 2 has the least number of NE areas, a high number of PE areas, a small volume of enhancing and T_2 hyperintense area, the most ellipsoid shape of enhancing and T_2 hyperintense areas.

Considering perfusion markers, subgroup 2 has mostly hypo-perfused DSC in PE areas and T_2 areas. Considering ASL perfusion, NE and PE areas are iso-perfused.

Subgroup 3

With regards to structural markers, subgroup 3 has relatively high number of NE areas, relatively the highest number of PE areas, the largest volume of enhancing and T_2 hyperintense areas and medium sized enhancing and T_2 area and shape. Considering perfusion markers, subgroup 3 has relatively high amount of DSC hyper-perfusion in NE areas, it has the highest amount of hyper-perfusion in PE areas and shows relatively the most hyper-perfused but mostly iso-perfused T_2 areas. Considering ASL perfusion, subgroup 3 has the highest amount of PE areas with hyper-perfusion and . Lastly, relatively highest amount of hyper-perfusion in T_2 areas.

Subgroup 4

Considering structural markers, subgroup 4 has a high number of NE and PE areas, the largest volume of enhancing and T_2 hyperintense areas and the most round shape of enhancing and T_2 areas. Considering DSC perfusion markers, subgroup 4 has the highest amount of hyper-perfused NE areas, has mostly hypo-perfused PE areas and shows mostly hypo-perfused T_2 areas. Concerning ASL perfusion markers, subgroup 4 has the highest number of hyper-perfused NE areas, a relatively high amount of PE areas with hyper-perfusion and some T_2 areas with hyper-perfusion.

As part of the survival analysis we performed Cox regression analysis and looked at the model when not corrected for clinical confounding variables. This allowed us to have a better insight of what the correction would do to our results. The results represent the association between different MRI phenotypes of glioblastomas and progression and survival.

In the uncorrected model, subgroup 3 has a significantly increased mortality risk (HR: 2.4 (CI: 1.1 – 5.0); p=0.03) compared to subgroup 2 (Supplementary Table S3; Supplementary Figure S2).

7.9.3 Supplementary tables

Supplementary Table S1. Clinical characteristics per subgroup.
The data represents mean ± SD, n (percentage) or median (interquartile range). Statistical analyses were performed with chi-square tests.

	Subgroup 1 (n=12)	Subgroup 2 (n=13)	Subgroup 3 (n=17)	Subgroup 4 (n=25)	p-value
Age	54 ± 12	56 ± 15	68 ± 11	60 ± 12	0.013
Female	4 (33%)	4 (31%)	8 (47%)	10 (40%)	0.799
IDH					
Wild-type	10 (83%)	6 (46%)	13 (76%)	19 (76%)	0.243
Mutant	0 (0%)	3 (23%)	1 (6%)	1 (4%)	
Not otherwise specified	2 (17%)	4 (31%)	3 (18%)	5 (20%)	
MGMT status					
Positive	2 (17%)	4 (31%)	6 (35%)	7 (28%)	0.741
Negative	10 (83%)	9 (69%)	11 (65%)	18 (72%)	
KPS median	90 (30)	90 (40)	90 (30)	90 (40)	
KPS 100	5 (42%)	2 (15%)	4 (24%)	7 (28%)	0.365
KPS 90	5 (42%)	5 (38%)	5 (29%)	8 (32%)	
KPS 80	1 (8%)	3 (23%)	5 (29%)	4 (16%)	
KPS 70	1 (8%)	2 (15%)	3 (18%)	3 (12%)	
KPS ≤60	0 (0%)	1 (8%)	0 (0%)	3 (12%)	
Surgery type					
Partial resection	2 (17%)	5 (38%)	9 (53%)	15 (60%)	<0.001
Total resection	9 (75%)	8 (62%)	2 (12%)	3 (12%)	
Biopsy	1 (8%)	0 (0%)	6 (35%)	7 (28%)	
Radiotherapy (total dose (Gy))					
40	1 (8%)	1 (8%)	4 (24%)	5 (20%)	0.458
45	2 (17%)	1 (8%)	3 (18%)	3 (12%)	
60	9 (75%)	11 (84%)	9 (53%)	17 (68%)	
Temozolomide chemotherapy	11 (92%)	10 (77%)	13 (77%)	13 (77%)	0.705

Supplementary Table S2. MRI markers of the glioblastomas per subgroup.
The data represents mean ± SD or n (percentage). Statistical analysis was performed with a chi-square test for categorical variables and a one-way ANOVA for the continuous variables. Most markers, except for most of the tumor location markers, and DSC perfusion in T_2 areas, differed significantly between subgroups ($p<0.05$).

Ordinal variables	**Subgroup 1 (n=12)**	**Subgroup 2 (n=13)**	**Subgroup 3 (n=17)**	**Subgroup 4 (n=25)**	**p-value**
Frontal (N(%))	2 (17%)	6 (46%)	8 (32%)	8 (32%)	0.11
Temporal (N(%))	6 (50%)	2 (15%)	0 (0%)	0 (0%)	0.01
Occipital (N(%))	0 (0%)	3 (23%)	2 (12%)	3 (12%)	0.37
Parietal (N(%))	2 (17%)	2 (15%)	2 (12%)	5 (20%)	0.92
Parieto-temporal (N(%))	1 (8%)	0 (0%)	1 (4%)	1 (4%)	0.77
Basal-nuclei (N(%))	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0.40
Corpus-callosum (N(%))	0 (0%)	0 (0%)	2 (12%)	1 (4%)	0.35
Insular (N(%))	0 (0%)	0 (0%)	1 (4%)	1 (4%)	0.72
Pons (N(%))	1 (8%)	0 (0%)	0 (0%)	0 (0%)	0.12
Presence of a nodular enhancing lesion (N(%))	11 (92%)	7 (54%)	15 (88%)	22 (88%)	0.03
Presence of a patchy enhancing lesion (N(%))	3 (25%)	12 (92%)	17 (100%)	24 (96%)	<0.001
Number of nodular lesions					
No lesions (N(%))	1 (8%)	7 (54%)	1 (6%)	3 (12%)	0.008
One lesion (N(%))	11 (92%)	6 (46%)	12 (71%)	16 (64%)	
Two lesions (N(%))	0 (0%)	0 (0%)	3 (18%)	6 (24%)	
Three lesions (N(%))	0 (0%)	0 (0%)	1 (6%)	0 (0%)	
Number of patchy lesions					
No lesions (N(%))	8 (67%)	0 (0%)	0 (0%)	0 (0%)	<0.001
One lesion (N(%))	4 (33%)	12 (92%)	13 (77%)	19 (76%)	
Two lesions (N(%))	0 (0%)	1 (8%)	2 (12%)	6 (24%)	
Three lesions (N(%))	0 (0%)	0 (0%)	2 (12%)	0 (0%)	
DSC perfusion in nodular lesions					
Hypoperfused (N(%))	7 (58%)	1 (8%)	2 (12%)	4 (16%)	0.004
Isoperfused (N(%))	4 (33%)	11 (85%)	11 (65%)	11 (44%)	
Hyperperfused (N(%))	1 (8%)	1 (8%)	4 (24%)	10 (40%)	

DSC perfusion in patchy lesions					
Hypoperfused (N(%))	0 (0%)	12 (92%)	3 (18%)	23 (92%)	<0.001
Isoperfused (N(%))	11 (92%)	1 (8%)	1 (6%)	2 (8%)	
Hyperperfused (N(%))	1 (8%)	0 (0%)	13 (77%)	0 (0%)	

DSC perfusion in T_2 lesions					
Hypoperfused (N(%))	1 (8%)	1 (8%)	1 (6%)	1 (4%)	0.73
Isoperfused (N(%))	11 (92%)	12 (92%)	14 (82%)	23 (92%)	
Hyperperfused (N(%))	0 (0%)	0 (0%)	2 (12%)	1 (4%)	

ASL perfusion in nodular lesions					
Isoperfused (N(%))	10 (83%)	13 (100%)	11 (65%)	13 (52%)	0.01
Hyperperfused (N(%))	2 (17%)	0 (0%)	6 (35%)	12 (48%)	

ASL perfusion in patchy lesions					
Isoperfused (N(%))	12 (100%)	13 (100%)	1 (6%)	20 (80%)	<0.001
Hyperperfused (N(%))	0 (0%)	0 (0%)	16 (94%)	5 (20%)	

ASL perfusion in T_2 lesions					
Isoperfused (N(%))	12 (100%)	13 (100%)	11 (65%)	23 (92%)	0.005
Hyperperfused (N(%))	0 (0%)	0 (0%)	6 (35%)	2 (8%)	

Continuous markers	Mean	SD	Mean	SD	Mean	SD	Mean	SD	p-value
Volume of enhancing lesions (ml)	19.49	29.37	7.14	5.21	247.54	249.60	374.45	282.52	<0.001
Volume of T2 lesions (ml)	329.44	338.37	181.19	148.76	1321.70	1094.77	1375.43	739.15	0.02
Eccentricity of enhancing lesions	0.65	0.22	0.80	0.15	0.68	0.13	0.60	0.18	<0.001
Eccentricity of T2 lesions	0.63	0.18	0.78	0.11	0.74	0.12	0.73	0.11	0.04

Supplementary Table S3. Regression analysis results for both adjusted and unadjusted models where the probability of survival of the subgroups is compared to subgroup 2, the subgroup with the least aggressive MRI markers.

		Model 1	Model 2
	Survival[a] (median (range))	Hazard ratio (95%-CI)	Hazard ratio (95%-CI)
Subgroup 1	13 (10 – 21)	1.7 (0.7 – 3.8)	2.6* (1.1 – 6.3)
Subgroup 2	22 (15 – 29)	Reference group	
Subgroup 3	11 (7 – 14)	2.4* (1.1 – 5.0)	1.4 (0.6 – 3.2)
Subgroup 4	10 (8 – 18)	1.5 (0.8 – 3.1)	0.995 (0.5 – 2.1)

[a] in months, * Significant (p-value < 0.05); CI: confidence interval.

The subgroup with the least aggressive MRI markers of glioblastomas (subgroup 2) was used as reference for the models. Model 1 was unadjusted and model 2 adjusted as multivariable analysis correcting for age, KPS, and extent of tumor surgical resection.

Supplementary Table S4. Regression analysis results for both adjusted and unadjusted models where the probability of progression of the subgroups is compared to subgroup 2, one of the subgroups with the least aggressive MRI markers.

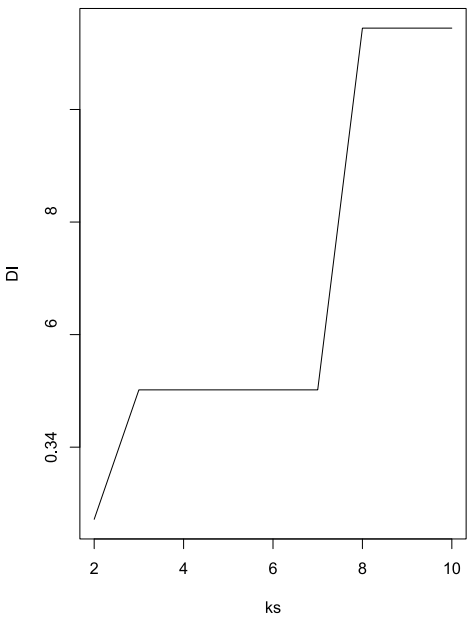
		Model 1	Model 2
	True progression (N(%))	Hazard ratio (95%-CI)	Hazard ratio (95%-CI)
Subgroup 1	67%	1.8 (0.4 – 8.3)	1.3 (0.2 – 6.8)
Subgroup 2	69%	Reference group	
Subgroup 3	82%	0.8 (0.8 – 3.4)	1.7 (0.3 – 10.4)
Subgroup 4	72%	0.9 (0.2 – 3.8)	1.7 (0.3 – 8.4)

[a] in months, * Significant (p-value < 0.05); CI: confidence interval.
The subgroup with the least aggressive MRI markers of glioblastomas (subgroup 2) was used as reference for the models. Model 1 was unadjusted and model 2 adjusted as multivariate analysis correcting for age, KPS, and extent of tumor surgical resection.

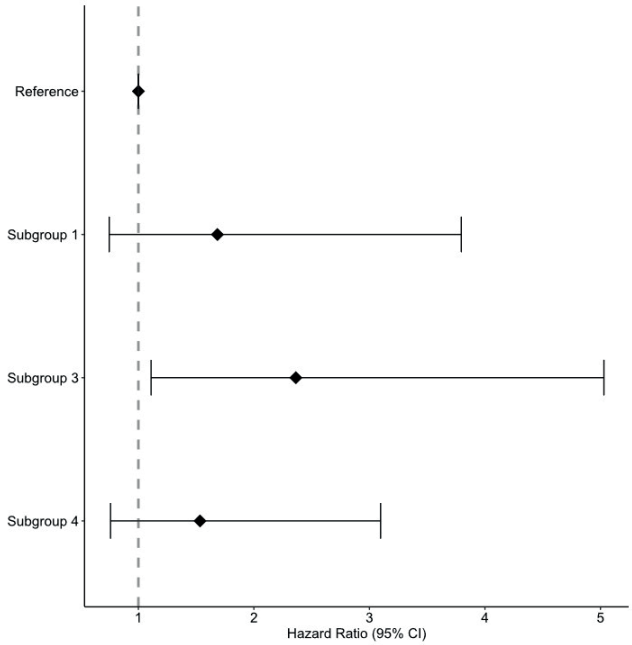
Supplementary Table S5. Log-rank test results comparing the overall survival times between subgroups and the reference group, subgroup 2, the subgroups with the least aggressive MRI markers.

Subgroups	Medial survival time in months (range)	X^2	p-value
Subgroup 1	13 (10 – 21)	1.919	0.166
Subgroup 2	11 (7 – 14)	6.906	0.009
Subgroup 3	10 (8 – 18)	1.710	0.191

7.9.4 Supplementary figures



Supplementary Figure S1. The Dunn index of the hierarchical clustering model. The y-axis shows the Dunn Index (DI) and the x-axis the number of clusters (ks).



Supplementary Figure S2. Forrest plot of the hazard ratios (95%-CI) per subgroup for survival.
Illustrated results of the cox proportional hazards survival analysis are shown unadjusted for confounding clinical variables. Subgroup 2 has been set as a reference and marked with the striped line in red. Hazard ratios are shown with a 95% confidence interval.

Supplementary Figure S3. Survival curves resulting from log-rank test between subgroups.
The results from a log-rank test are shown illustrating how the median overall survival time differs between the reference subgroup (subgroup 2) and subgroup 1 (A), subgroup 3 (B) and subgroup 4 (C).