

Quantifying eloquent locations for glioblastoma surgery using resection probability maps

Abstract

Object. Decisions in glioblastoma surgery are often guided by presumed eloquence of the tumor location. We introduce the ‘expected residual tumor volume’ (eRV) and the ‘expected resectability index’ (eRI) based on previous decisions aggregated in resection probability maps. We determine the diagnostic accuracy of eRV and eRI to predict biopsy decisions, resectability, functional outcome and survival.

Methods. Consecutive patients with first-time glioblastoma surgery in 2012-2013 were included from 12 hospitals. We calculated the eRV from the preoperative MR-scans of each patient using a resection probability map, and derived the eRI from the tumor volume. As reference, we classified Sawaya’s eloquence grades (EG). Resectability was measured as observed extent of resection and residual volume, functional outcome as change in Karnofsky performance. Receiver operating characteristics curves and multivariable regression were applied.

Results. Of 915 patients, 674 (74%) had a resection with a median extent of resection of 97%, functional decline in 71 (8%), and median survival of 12.8 months. The eRI and eRV identified biopsies and extents of resection of at least 80%, 90% or 98% better than EG. The eRV and eRI predicted observed residual volumes under 10, 5 and 1 mL better than EG. The eRV, eRI and EG had low diagnostic accuracy for functional outcome changes. Higher eRV and lower eRI were strongly associated with shorter survival, independent of known prognostic factors.

Conclusions. The eRV and the eRI predict biopsy decisions, resectability and survival better than eloquence grading and may be useful preoperative indices to support surgical decisions.

Keywords

glioma; neurosurgery; extent of resection; residual volume; reproducibility of results

Introduction

Surgical decisions in patients with glioblastoma vary from no treatment to biopsy, partial resection, radiologically complete resection or so-called supratotal resection. The choice is guided by the aim to maximize tumor removal to prolong survival on one side and the aim to minimize the risk of permanent functional deficits on the other. The main arguments for these surgical decisions include the resectability of the tumor and the expected impact on survival and functional performance. Patient age and condition are typically integrated in this decision-making, as well as the tumor volume and location.

Resectability has been expressed as percentage resectable volume of the preoperative tumor volume or as volume of residual tumor, and notable disagreement has been reported between expected and observed resectability.^{4,9,20,39,40} Reports have estimated a minimum threshold for extent of resection between 78% and 98%, and a maximum residual volume between 1 and 5 mL to prolong survival of glioblastoma.^{7,32} Instead of a threshold, others have reported a continuous positive relationship between resection results and survival.^{15,27}

Brain regions to avoid during surgery are sometimes referred to as ‘eloquent’, i.e. functionally of critical importance, which will result in loss of brain functions, if removed or damaged. Nowadays intraoperative stimulation mapping is the standard to identify these brain regions in individual patients.^{12,13} It helps however, to have a measure of brain function prior to surgery to guide decisions on biopsies and whether and where in the brain to apply intraoperative stimulation mapping. Several methods identify functions non-invasively. Some methods are based on population-based classification of structural anatomy or atlases of brain functions. Sawaya’s classification of eloquence grade (EG) has often been used in reports of surgical cohorts and was shown to correlate with postoperative neurological deficits.³⁴ Other surgical classifications exist for the whole brain^{4,36} or brain regions.²² Brain atlases have been

constructed for brain plasticity²¹ and cortical⁴² and subcortical³³ structures. Other methods are based on localizing brain function in individuals, such as preoperative diagnostic imaging with structural MRI⁴¹, task-based functional MRI⁵, resting state functional MRI¹⁰, magnetoencephalography¹, white matter tractography², or transcranial magnetic stimulation.²⁴

Another source of potentially useful information before surgery are resection probability maps based on a large number of prior resections. Resection probability maps of non-enhancing glioma have been previously used to estimate the resection result,^{20,26} to evaluate the potential for brain plasticity,²¹ and to compare resection results between surgical teams.^{11,29} A new application for resection probability maps is to quantify resectability for a new patient's tumor indicating the expected residual tumor portion.

In this study, we determined the diagnostic accuracy of the 'expected residual tumor volume' (eRV) and the 'expected resectability index' (eRI) as preoperative measures to guide biopsy decisions, to estimate resectability, and to predict functional outcome and survival in comparison with EG.

Methods

Patients

We identified all patients of at least 18 years old with a newly-diagnosed glioblastoma at first-time surgery between 1/1/2012 and 12/31/2013 from 12 hospitals: **[BLINDED FOR REVIEW #1-#12]**. Patients were identified at each hospital by prospective electronic databases and included if they complied with these inclusion criteria. Patients were excluded from analysis if they had non-enhancing glioblastoma on MRI, if an MR-scan was made later than 3 days postoperative⁴⁵, or if the imaging was of insufficient quality for segmentation and registration, due to missing images or severe motion artefacts. Patients received standard care and treatment decisions were made in multidisciplinary tumor board meetings.

From the medical records of these patients we extracted age at diagnosis, gender, type of surgery, the Karnofsky performance preoperative and at two months postoperative, the dates of surgery, last follow-up in the clinic, and death. We also obtained the MR-scans pre- and postoperative.

Imaging protocols were standardized in hospitals although not identical between hospitals and included 3D heavily T1-weighted gradient-echo pulse sequence at 1 mm isotropic resolution, at most 5 mm, obtained after administration of intravenous gadolinium, a T2/FLAIR-weighted gradient-echo pulse sequence and a diffusion-weighted sequence.

Approval of the study protocol was obtained by the institutional review boards and informed consent from patients was obtained according to local regulations. Data and imaging for analysis were anonymized.

Expected residual tumor volume, expected resectability index, and eloquence grade

Surgical eloquence was quantitated using a resection probability map derived from all patients. For details on the methodology see Supplemental Figure 1 and our earlier description.²⁹ This resection probability map represents probability of resection at 1 mm³ resolution in standard brain space, i.e. the MNI-152 brain template,¹⁷ based on pre- and postoperative tumor segmentations.^{11,26,29} We used software (Smartbrush Suite Software; BrainLAB AG, Munich, Germany) to outline the contrast-enhancing elements on the T1 images with enclosed necrosis or cyst in three orthogonal planes. For postoperative residual tumor interpretation, we inspected T1-weighted series without gadolinium and diffusion-weighted series to distinguish postsurgical artefacts from residual tumor in patients who had had a resection. In patients undergoing biopsy the postoperative residual tumor was considered identical to the preoperative tumor segmentation. To calculate the resectability, we used the patient's preoperative tumor segmentation to mask the resection probability map derived from all other patients by a leave-one-out approach. The resection probabilities of the masked voxels for that patient were then integrated to obtain the expected resectable volume. The preoperative tumor volume minus the expected resectable volume resulted in the expected residual tumor volume, eRV, expressed in mL.²⁰ A division of the expected resectable volume by the preoperative tumor volume resulted in the expected resectability index, eRI, ranging between 0.0 and 1.0 on a continuous scale. A value of 0.0 represented a non-resectable tumor and a value of 1.0 a completely resectable tumor. A webapplication was made to enable calculations of the eRV and eRI from preoperative MRIs of new patients (<https://www.pictureproject.nl>).

In addition, we classified the eloquence grade (EG) for each patient's tumor ranging from noneloquent (grade I), near-eloquent (grade II) to eloquent (grade III) as a historical

reference.³⁴ For eloquence grading we used an automated approach based on the registered preoperative tumor volumes, which is therefore completely reproducible and devoid of subjective prior knowledge of preoperative neurological symptoms, presumed language dominance or resection outcomes. For this, we determined for each patient the number of tumor voxels overlapping with the regions corresponding to these eloquence grade definitions³⁴, as indicated in Figure 1E. The majority vote of voxel summaries determined the eloquence grade.

Outcome measures

We considered four outcome measures in their relation with eRV, eRI and EG: a biopsy decision, the observed extent of resection, postoperative functional changes and survival.

For the biopsy decisions, we evaluated their association with eRV, eRI and EG in conjunction with a number of predictive factors, i.e. age, gender, preoperative Karnofsky performance, and tumor volume. For the observed extents of resection, we calculated the postoperative residual volume and preoperative tumor volume. The percentage of the preoperative volume, that had been resected, was considered the extent of resection. We evaluated the eRV, eRI and EG in their prediction of a range of extents of resection of over 80%, 90% and 98%, and a range of observed residual tumor volumes of less than 10, 5 and 1 mL. For the postoperative Karnofsky performance changes, we subtracted the performance at two months postoperative from the preoperative performance. An increase in KPS of 20 points or more was considered a functional improvement; a decrease in KPS of 20 or less was considered a functional decline.⁶ We related postoperative performance change with the eRV, eRI and EG together with age, gender, preoperative performance and tumor volume. For the survival time, we calculated the time between diagnosis and death, or the patient was considered censored at the last date of follow-up. We correlated survival with the eRV, eRI

and EG in conjunction with the established prognostic factors: age and preoperative performance, as well as with gender and tumor volume.

Statistical analysis

The correlation between eRV and EG and eRI and EG was determined by Kendall's tau.

To determine the diagnostic accuracy of eloquence quantifications for outcome measures, we used receiver operating characteristics plots. We calculated the areas under the curve (AUC) of sensitivity versus specificity and compared AUC's using the bootstrap method³¹ and interpreted as poor, reasonable, good, and excellent.³⁷ P-values less than 0.05 were considered significant. Predictions that included clinical information, such as age, performance, and tumor volume were calculated with the regression coefficients from the multivariable models.

We used multivariable logistic regression with biopsy as binary response variable and eRV, eRI, EG, age, gender, preoperative performance and tumor volume as dependent variables. The regression assumptions were visually verified and met.

We used survival analysis with survival time as response variable and eRV, eRI and EG as predictors and age, gender, preoperative performance and tumor volume as confounders in multivariable proportional hazards regression model. Kaplan-Meier curves were drawn for low, intermediate and high eRV and eRI and for EG. Residuals were verified and model assumptions were met.

Results

A total of 1083 patients were identified and evaluated for inclusion. Of these, 110 patients were excluded because the MR-scan was later than three days postoperative, 53 patients had MR-scans of insufficient quality to register to standard space, and five patients had tumors not enhancing with contrast. The characteristics of the 915 patients for analysis are listed in Supplemental Table 1. Patient characteristics were comparable across teams.

The tumor probability map, the resection probability map and the EG are shown in Figure 1. The EG does not correlate with the eRV (Kendall's tau: 0.01, Figure 1F), nor with the eRI (Kendall's tau: 0.02, Figure 1G). The eRV and the eRI are correlated (Kendall's tau: -0.28, $p < 0.00001$, Figure 1H). Three examples of decreasing resectability are demonstrated in Figure 2A-C.

Biopsy indications and resectability

Overall, 241 (26%) patients were biopsied and 674 (74%) patients had a resection with a median extent of resection of 97% (Supplemental Table 1).

To evaluate the relation between eRV, eRI and EG to predict resectability, we first determined the diagnostic accuracy to identify patients who were biopsied (Figure 3A). The eRI identified biopsies significantly better (AUC, 95%CI: 0.77, 0.74-0.81, $p < 0.00001$) than the eRV (0.52, 0.48-0.56) and the EG (0.47, 0.43-0.51). Although the discrimination using eRI is good, a positive predictive value was observed of 48%, and a negative predictive value of 87% at the optimal threshold of 0.72 that maximizes sensitivity (71%) and specificity (72%). Second, we determined the accuracy to predict extent of resection of at least 98%, 90% and 80% for patients with a resection (Figure 3B). The eRI was significantly better to identify resections at least 98% (0.61, 0.57-0.65, $p < 0.00001$), 90% (0.65, 0.61-0.69, $p < 0.00001$) and 80% (0.66, 0.60-0.71, $p = 0.00005$) and the eRV (0.60, 0.55-0.63, $p = 0.0003$;

0.58, 0.53-0.63, $p=0.005$; and 0.56, 0.50-0.62, n.s., respectively) than the EG (0.49, 0.45-0.53; 0.50, 0.45-0.54; and 0.48, 0.43-0.54, respectively). Again, the discrimination using eRI to identify patients with an extent of resection at least 98% is sufficient, with a positive predictive value of 64%, and a negative predictive value of 56% at the optimal threshold of 0.80 that maximizes sensitivity (69%) and specificity (51%). Third, we examined the accuracy to predict residual volume of less than 10, 5 and 1 mL for patients with a resection (Figure 3C). The eRV was significantly better to identify observed residual volumes less than 10 mL (0.81, 0.77-0.85, $p<0.00001$), 5 ml (0.77, 0.73-0.81, $p<0.00001$) and 1 mL (0.73, 0.69-0.76, $p<0.00001$) and the eRI (0.66, 0.60-0.72, $p=0.003$; 0.65, 0.61-0.70, $p<0.00001$; and 0.63, 0.59-0.67, $p<0.00001$, respectively) than the EG (0.53, 0.46-0.59; 0.49, 0.44-0.54; and 0.49, 0.45-0.54, respectively). The discrimination using eRV to identify patients with a residual volume less than 1 mL is good with a positive predictive value of 62%, and a negative predictive value of 77% at the optimal threshold of 4.6 that maximizes sensitivity (84%) and specificity (51%).

Clearly, the eRV and eRI are imperfect to identify the biopsy indications (Figure 3A). To improve this with clinical information available at the time of the decision, we further explored the association between the eRV and eRI and the biopsy decisions in conjunction with age, performance, and tumor volume (Supplemental Table 2). Higher eRV, lower eRI, and increasing EG had a strong association with the decision to biopsy in conjunction with age, performance and tumor volume. The biopsy predictions based on the eRI were significantly improved in combination with the available clinical information (Figure 3A; AUC: 0.84, 0.82-0.87, $p=0.0022$).

The eRV and eRI are also imperfect to identify the patients with resections over 98% (Figure 3B). Lower eRV, higher eRI and decreasing EG had a strong association with resections over 98% in conjunction with age, performance and tumor volume (Supplemental

Table 2). Resections over 98% based on the eRI were significantly improved by combining the clinical variables (Figure 3B; AUC: 0.68, 0.64-0.72, $p=0.019$). Furthermore the eRV was better than eRI to identify resections with less than 1 mL residual volume, which could not be improved further with the clinical variables (Figure 3C; Supplemental Table 2).

Functional outcome

Overall 71 (8%) patients improved functionally and 78 (9%) declined between surgery and two months follow-up. Functional change information was missing in 90 (10%) patients.

The eRV, eRI and EG were not related with changes in performance by visual inspection, although the extremes of change, both positive and negative, were indicated to occur at low eRV and high eRI (Figure 4A-C). An increased odds of functional decline was significantly related with a lower eRI in logistic regression, as well as with higher preoperative Karnofsky performance and higher tumor volume (Supplemental Table 3). The eRV and EG were not associated with functional decline in logistic regression. An increased odds of functional improvement was significantly associated with lower eRV and higher eRI, as well as with lower age, lower performance and the decision to resect (Supplemental Table 3).

Apparently, the decision to resect was associated with functional improvement and not with functional decline. Expanding on this, a measure of surgical aggressiveness could be constructed by subtracting the observed from the expected residual tumor volume and similarly subtracting the resectability index from the observed extent of resection. These measures of surgical aggressiveness clearly separate biopsies from resections (Figure 4D-E), but do not indicate surgical aggressiveness to be associated with functional changes.

The diagnostic accuracy of eRI to identify functional decline (Figure 5A) was significantly better (AUC, 95%CI: 0.60, 0.53-0.67, $p=0.026$) than the EG (0.50, 0.43-0.56), whereas eRV was not (0.59, 0.52-0.67, $p=0.052$). Although the discrimination using eRI is

poor with a positive predictive value of 11%, and a negative predictive value of 95% at the optimal threshold of 0.75 that maximizes sensitivity (68%) and specificity (52%). The diagnostic accuracy of eRV, eRI and EG to identify functional improvement (Figure 5B) were similarly poor (0.57, 0.50-0.63; 0.58, 0.51-0.64; and 0.51, 0.44-0.58, respectively).

Adding the available clinical information age, performance, tumor volume and resection decision significantly improved the identification of functional improvement (0.89, 0.85-0.92, $p < 0.00001$; Figure 5B) and to a lesser extent the identification of functional decline (0.70, 0.64-0.76, $p = 0.036$; Figure 5A).

Survival

The median overall survival is 12.8 months for the whole population.

To evaluate eRV, eRI and EG in their association with survival, we first plotted Kaplan-Meier curves (Figure 6). The eRV was significantly associated with survival (Figure 6A, $p < 0.00001$) with median overall survival times of 16, 13 and 9 months, respectively, for low, intermediate and high eRV at cutoffs of 1.7 and 17 mL. The eRI was also significantly associated with survival (Figure 6B, $p < 0.00001$) with median overall survival times of 16, 13 and 8 months, respectively, for high, intermediate and low eRI at cutoffs of 0.8 and 0.5. The EG was not associated with survival (Figure 6C, n.s.). For comparison, actual biopsy decision and observed extent of resection - only available after surgery - were confirmed to be strongly associated with survival (Fig 6D, $p < 0.00001$) with median overall survival time of 18, 13 and 6 months for extents of resection over 98%, under 98% and biopsies.

To determine the interaction between the eRV and eRI and known prognostic factors, we further explored this association in multivariable models (Supplemental Table 4). A larger eRV and a lower eRI were significantly associated with shorter survival, independent from higher age, lower performance, and larger tumor volume. EG was not associated with

survival. Lesser observed extent of resection was confirmed to be significantly associated with shorter survival, independent of the known prognostic factors.

Discussion

Surgical eloquence can be quantified as the ‘expected residual tumor volume’ and the ‘expected resectability index’ based on previous surgical results in other patients. These measures are potentially helpful to guide surgical decision-making, because they are associated with biopsy decisions, the observed residual tumor volume and the extent of resection. The eRV and eRI do not discern patients prone to functional decline or improvement, but are a suitable preoperative surrogate marker for survival. These measures also enable the quantification of surgical aggressiveness as demonstrated. The implication of our findings is that structured previous surgical results contain quantitative information to better inform new surgical decisions.

Resection probability maps, from which the eRV and eRI are derived, can be valued as an aggregated snapshot of how neurosurgeons perceive functional and otherwise critical regions of the brain. Resection probability maps can be used in combination with other methods, such as brain atlases of positive stimulation sites³³, to guide surgical decisions. Importantly, the noninvasive eRV and eRI are available preoperatively to guide treatment decisions, whereas the standard for location of brain functions is the identification with electrostimulation during surgery. Resectability, functional outcome and survival have been related with the positive stimulation areas^{9,12} as well as with eloquence grading.^{7,25,32,38} For patients who have surgery without stimulation mapping, the ‘eloquence’ of brain regions in relation to the tumor will only be based on noninvasive preoperative information, which should therefore be optimized.

The eRV and eRI are probably most useful for clinical decisions before surgery. Surgical decisions are now based on intuitive estimates of resectability with input from presumed functionality of tumor-infiltrated brain regions with associated risk for functional deficits and

from a presumed increase in life expectancy by removing more tumor tissue.²⁷ The eRV and eRI seem to improve the EG, that, although mainly of historical importance, is still a common basis for presumed functionality. More informative than a single metric are the mappings of previous surgical results on the preoperative structural MRI of a new patient, as we have now demonstrated. This information can provide additional arguments to determine whether and where to use stimulation mapping during resections.

The predictions of biopsy decisions, resectability, functional decline and improvement and survival are not very accurate, and strict thresholds for eRV and eRI should probably not be applied in clinical practice. Notwithstanding the less than optimal accuracy, these predictions may improve current practice, given the large variation in rates of biopsies and gross total resections between reports of large populations,^{3,19,35} which suggests considerable treatment variation and outcome variation. The eRV and eRI are strongly associated with biopsy decisions, resectability, functional decline and survival compared with age, performance and tumor volume. Given the imperfect predictions, other factors are likely involved as well.^{27,34,39} An explanation for imperfect estimation of resectability in this population may be heterogeneity among neurosurgeons to biopsy or to apply surgical techniques, such as image guidance² and intraoperative stimulation mapping.¹² For instance, several neurosurgical teams have replaced biopsies by resections with intraoperative stimulation mapping, supported by the widely varying biopsy percentages in our population. An explanation for imperfect prediction of functional decline may be that decline not only results from surgical removal of functional brain, but also from vascular injury, medical complications or early disease progression.¹⁸ Other factors that may be involved are type and severity of symptoms and symptom relief on steroids. Furthermore, an explanation for imperfect prediction of survival in this population may be differences in treatment guideline adherence, adjuvant treatment concessions, clinical trial participation, endurance in treating

progression, molecular markers, tumor growth characteristics and the patient's preferences at the end of life.

Several strategies may further increase the accuracy of the resectability quantifications in future studies. For instance, a two-step approach could be devised to first separate the patients for biopsy and then to estimate the resectability of the remaining patients for resections based on others who have had a resection. Consensus biopsy indications could be determined for the resection probability map population by an expert panel. Alternatively, a classifier for biopsy indications could be constructed based on a range of patient characteristics, including the initial MRI, using machine learning techniques. Furthermore, the resectability prediction for an individual patient could be based on a resection probability map from similar patients in terms of age, condition, location, size and aspect of the tumor. Ideally the map would be iterative and incorporate data from additional patients to further improve decisions based on better predictions. Our finding, that surgical aggressiveness - as difference between expected and observed tumor removal - is not associated with functional changes, supports that neurosurgeons succeed to avoid functional decline from overextensive resections into essential functional brain structures. Conversely, this may also indicate that the more aggressive end of the resection spectrum is nevertheless more conservative than necessary, because erring on the side of less tumor removal is generally preferred over too extensive tumor removal.

A basic assumption of our study is that the location of the tumor determines the surgical strategy. And consequently, if the surgical options are limited, then survival will be shorter. However, another perspective is that the location of the tumor in itself is related to molecular subtypes.^{14,43} For instance, a patient with an eloquently-located tumor of a more favorable subtype could benefit from aggressive maximization of the resection, whereas a patient with a noneloquently-located tumor of a very unfavorable subtype may be better off with a biopsy

only. As an example, glioblastoma near lateral ventricular zones⁸ and white matter intersections²⁸ have been associated with poor survival despite maximal resections.

Strengths of this study are that well-documented complete patient populations from 12 neuro-oncological referral hospitals throughout Europe and North America were combined to represent the heterogeneity of current neurosurgery. The sizeable population enabled the construction of a high-resolution resection probability map that can be extrapolated to other neurosurgical practices. The voxel-based approach to residual tumor information avoided assumptions on anatomical classifications for the aggregation of results. We have used state-of-the-art methods for tumor segmentation and brain registration to standard space.

Limitations of this study are that the residual tumors that were used for the probability map were not necessarily restricted to intentional residual tumor to preserve functionality. An unknown portion of residual tumors may have been overlooked during surgery. This effect could blur the image of brain functionality according to neurosurgeons. We adhered to the standards¹⁶ by segmenting the gadolinium-enhancing portions of glioblastoma, which is known to undersample the true extent of glioblastoma.²³ In addition, neuroimaging protocols, manual tumor segmentation⁴⁴ and brain registration³⁰ can be subject to variation and are therefore potential sources of error, although visual inspection verified satisfactory processing in all patients. The Karnofsky performance is a coarse measure of functional status that may have missed neurological and cognitive changes.

Conclusions

Previous surgical results can be quantified into the eRV and eRI as preoperative measures that predict biopsy decisions, resectability, and survival better than eloquence grading.

Acknowledgements

We would like to thank Dr Emmanuel Mandonnet for kindly providing the patient data from Hôpital Lariboisière in Paris and commenting on the manuscript. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative and the Translational Research IT (TraIT) project, an initiative from the Center for Translational Molecular Medicine (CTMM). We thank the QuantiVision institute for collaborating in applying for the Netherlands Organisation for Scientific Research grant program.

Disclosure

Brainlab provided segmentation software as a contribution in kind to this study. The authors declare that they have no competing interests.

References

1. Baillet S: Magnetoencephalography for brain electrophysiology and imaging. **Nat Neurosci** **20**:327–339, 2017
2. Barone DG, Lawrie T a, Hart MG, Dg B, Ta L, Mg H: Image guided surgery for the resection of brain tumours. **Cochrane database Syst Rev** **1**:CD009685, 2014
3. Bauchet L, Mathieu-Daude H, Fabbro-Peray P, Rigau V, Fabbro M, Chinot O, et al: Oncological patterns of care and outcome for 952 patients with newly diagnosed glioblastoma in 2004. **Neuro Oncol** **12**:725–735, 2010
4. Capellades J, Puig J, Domenech S, Pujol T, Oleaga L, Camins A, et al: Is a pretreatment radiological staging system feasible for suggesting the optimal extent of resection and predicting prognosis in glioblastoma? An observational study. **J Neurooncol** **137**:367–377, 2018
5. Castellano A, Cirillo S, Bello L, Riva M, Falini A: Functional MRI for Surgery of Gliomas. **Curr Treat Options Neurol** **19**:34, 2017
6. Chaichana KL, Halthore AN, Parker SL, Olivi A, Weingart JD, Brem H, et al: Factors involved in maintaining prolonged functional independence following supratentorial glioblastoma resection. **J Neurosurg** **114**:604–612, 2011
7. Chaichana KL, Jusue-Torres I, Navarro-Ramirez R, Raza SM, Pascual-Gallego M, Ibrahim A, et al: Establishing percent resection and residual volume thresholds affecting survival and recurrence for patients with newly diagnosed intracranial glioblastoma. **Neuro Oncol** **16**:113–122, 2014
8. Chaichana KL, McGirt MJ, Frazier J, Attenello F, Guerrero-Cazares H, Quinones-Hinojosa A: Relationship of glioblastoma multiforme to the lateral ventricles predicts

- survival following tumor resection. **J Neurooncol** **89**:219–224, 2008
9. Coburger J, Renovanz M, Ganslandt O, Ringel F, Wirtz CR, von Riehm JS: Evaluation of surgical decision making and resulting outcome in patients with highly eloquent glioblastoma: Results of a multicenter assessment. **Clin Neurol Neurosurg** **162**:29–35, 2017
 10. Cochereau J, Deverdun J, Herbet G, Charroud C, Boyer A, Moritz-Gasser S, et al: Comparison between resting state fMRI networks and responsive cortical stimulations in glioma patients. **Hum Brain Mapp** **37**:3721–3732, 2016
 11. De Witt Hamer PC, Hendriks EJ, Mandonnet E, Barkhof F, Zwinderman AH, Duffau H: Resection Probability Maps for Quality Assessment of Glioma Surgery without Brain Location Bias. **PLoS One** **8**:2013
 12. De Witt Hamer PC, Robles SG, Zwinderman AH, Duffau H, Berger MS: Impact of intraoperative stimulation brain mapping on glioma surgery outcome: A meta-analysis. **J Clin Oncol** **30**:2559–2565, 2012
 13. Duffau H: Stimulation mapping of white matter tracts to study brain functional connectivity. **Nat Rev Neurol** **11**:255–265, 2015
 14. Ellingson BM, Lai A, Harris RJ, Selfridge JM, Yong WH, Das K, et al: Probabilistic radiographic atlas of glioblastoma phenotypes. **Am J Neuroradiol** **34**:533–540, 2013
 15. Ellingson BM, Abrey LE, Garcia J, Chinot O, Wick W, Saran F, et al: Post-chemoradiation volumetric response predicts survival in newly diagnosed glioblastoma treated with radiation, temozolomide, and bevacizumab or placebo. **Neuro-Oncology**:1–11, 2018
 16. Ellingson BM, Wen PY, Cloughesy TF: Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. **Neurotherapeutics** **14**:307–320, 2017

17. Fonov VS, Evans AC, McKinstry RC, Almlí CR, Collins DL: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. **Neuroimage** **47**:S102, 2009
18. Gempt J, Förschler A, Buchmann N, Pape H, Ryang Y-M, Krieg SM, et al: Postoperative ischemic changes following resection of newly diagnosed and recurrent gliomas and their clinical relevance. **J Neurosurg** **118**:801–808, 2013
19. Graus F, Bruna J, Pardo J, Escudero D, Vilas D, Barceló I, et al: Patterns of care and outcome for patients with glioblastoma diagnosed during 2008-2010 in Spain. **Neuro Oncol** **15**:797–805, 2013
20. Hendriks EJ, Habets EJJ, Taphoorn MJB, Douw L, Zwinderman AH, Vandertop WP, et al: Linking late cognitive outcome with glioma surgery location using resection cavity maps. **Hum Brain Mapp** **39**:2064–2074, 2018
21. Herbet G, Maheu M, Costi E, Lafargue G, Duffau H: Mapping neuroplastic potential in brain-damaged patients. **Brain** **139**:829–44, 2016
22. Hervey-Jumper SL, Li J, Osorio JA, Lau D, Molinaro AM, Benet A, et al: Surgical assessment of the insula. Part 2: validation of the Berger-Sanai zone classification system for predicting extent of glioma resection. **J Neurosurg**:1–7, 2015
23. Jain R, Poisson LM, Gutman D, Scarpace L, Hwang SN, Holder C a, et al: Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. **Radiology** **272**:484–93, 2014
24. Krieg SM, Lioumis P, Mäkelä JP, Wilenius J, Karhu J, Hannula H, et al: Protocol for motor and language mapping by navigated TMS in patients and healthy volunteers; workshop report. **Acta Neurochir (Wien)** **159**:1187–1195, 2017

25. Li YM, Suki D, Hess K, Sawaya R: The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? **J Neurosurg** **124**:977–988, 2016
26. Mandonnet E, Jbabdi S, Taillandier L, Galanaud D, Benali H, Capelle L, et al: Preoperative estimation of residual volume for WHO grade II glioma resected with intraoperative functional mapping. **Neuro Oncol** **9**:63–69, 2006
27. Marko NF, Weil RJ, Schroeder JL, Lang FF, Suki D, Sawaya RE: Extent of resection of glioblastoma revisited: Personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. **J Clin Oncol** **32**:774–782, 2014
28. Mickevicius NJ, Carle AB, Bluemel T, Santarriaga S, Schloemer F, Shumate D, et al: Location of brain tumor intersecting white matter tracts predicts patient prognosis. **J Neurooncol** **125**:393–400, 2015
29. Müller DMJ, Robe PAJT, Eijgelaar RS, Witte MG, Visser M, de Munck JC, et al: Comparing Glioblastoma Surgery Decisions Between Teams Using Brain Maps of Tumor Locations, Biopsies, and Resections. **JCO Clin Cancer Informatics**:1–12, 2019
30. Ou Y, Akbari H, Bilello M, Da X, Davatzikos C: Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights. **IEEE Trans Med Imaging** **33**:2039–2065, 2014
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al: pROC: an open-source package for R and S+ to analyze and compare ROC curves. **BMC Bioinformatics** **12**:77, 2011
32. Sanai N, Polley M-Y, McDermott MW, Parsa AT, Berger MS: An extent of resection

- threshold for newly diagnosed glioblastomas. **J Neurosurg** **115**:3–8, 2011
33. Sarubbo S, De Benedictis A, Merler S, Mandonnet E, Balbi S, Granieri E, et al: Towards a functional atlas of human white matter. **Hum Brain Mapp** **36**:3117–36, 2015
 34. Sawaya R, Hammoud M, Schoppa D, Hess KR, Wu SZ, Shi WM, et al: Neurosurgical outcomes in a modern series of 400 craniotomies for treatment of parenchymal tumors. **Neurosurgery** **42**:1044–1046, 1998
 35. Scoccianti S, Magrini SM, Ricardi U, Detti B, Buglione M, Sotti G, et al: Patterns of care and survival in a retrospective analysis of 1059 patients with glioblastoma multiforme treated between 2002 and 2007: A multicenter study by the central nervous system study group of Airo (Italian association of radiation oncology). **Neurosurgery** **67**:446–458, 2010
 36. Shinoda J, Sakai N, Murase S, Yano H, Matsuhisa T, Funakoshi T: Selection of eligible patients with supratentorial glioblastoma multiforme for gross total resection. **J Neurooncol** **52**:161–71, 2001
 37. Šimundić A-M: Measures of Diagnostic Accuracy: Basic Definitions. **EJIFCC** **19**:203–211, 2009
 38. Smith JS, Chang EF, Lamborn KR, Chang SM, Prados MD, Cha S, et al: Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas. **J Clin Oncol** **26**:1338–45, 2008
 39. Sonabend AM, Zacharia BE, Cloney MB, Sonabend A, Showers C, Ebiana V, et al: Defining Glioblastoma Resectability Through the Wisdom of the Crowd: A Proof-of-Principle Study. **Neurosurgery** **0**:2016
 40. Southwell DG, Birk HS, Han SJ, Li J, Sall JW, Berger MS: Resection of gliomas

- deemed inoperable by neurosurgeons based on preoperative imaging studies. **J Neurosurg** **129**:567–575, 2018
41. Talos I-F, Zou KH, Ohno-Machado L, Bhagwat JG, Kikinis R, Black PM, et al: Supratentorial low-grade glioma resectability: statistical predictive analysis based on anatomic MR features and tumor characteristics. **Radiology** **239**:506–13, 2006
 42. Tate MC, Herbet G, Moritz-Gasser S, Tate JE, Duffau H: Probabilistic map of critical functional regions of the human cerebral cortex: Broca’s area revisited. **Brain** **137**:2773–2782, 2014
 43. Tejada Neyra MA, Neuberger U, Reinhardt A, Brugnara G, Bonekamp D, Sill M, et al: Voxel-wise radiogenomic mapping of tumor location with key molecular alterations in patients with glioma. **Neuro Oncol** **20**:1517–1524, 2018
 44. Visser M, Müller DMJ, van Duijn RJM, Smits M, Verburg N, Hendriks EJ, et al: Inter-rater agreement in glioma segmentations on longitudinal MRI. **NeuroImage Clin** **22**:101727, 2019
 45. Wen PY, Macdonald DR, Reardon D a., Cloughesy TF, Sorensen a. G, Galanis E, et al: Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. **J Clin Oncol** **28**:1963–1972, 2010

Legends

Figure 1. The tumor probability maps for glioblastoma in the left (A, n=451) and in the right (B, n=464) hemisphere indicating the number of cases in the dataset involved by tumor within each voxel, the resection probability maps for the left (C) and right (D) hemisphere indicating the resectability index at 1 mm resolution and the eloquence grade (EG) (E) in standard brain space. The numbers across the bottom refer to Z-values of standard brain space. The correlation per patient between the EG and the expected residual tumor volume (F) and the resectability index (G), and between the expected residual tumor volume and the resectability index (H). The EG regions were specified by Sawaya.¹⁴ The color codes are according to the legends, where in A and B 95 is the maximum number of cases per voxel.

Figure 2. Case examples of decreasing resectability: (A) a 74-year-old male with a 13 mL tumor, an expected residual tumor volume (eRV) of 1.2 mL, a resectability index (eRI) of 0.92, an eloquence grade (EG) of I, an observed residual volume of 0 mL and 100% extent of resection (EOR), (B) a 63-year-old female with a 49 mL tumor, an eRV of 14 mL, a eRI of 0.70, an EG of II, an observed residual volume of 2.6 mL and an EOR of 95%, and (C) a 47-year-old male with a 40 mL tumor, an eRV of 22 mL, a eRI of 0.47, an EG of III, an observed residual volume of 7.0 mL and an EOR of 83%. Color coding is identical to Figure 1.

Figure 3. The diagnostic accuracies for (A) biopsy decisions, (B) extent of resection, and (C) and observed residual tumor volume by the expected residual tumor volume (eRV) in red, the expected resectability index (eRI) in green and the eloquence grade (EG) in blue as receiver operating characteristic curves. The grey curves incorporate age, performance and tumor volume in addition to the eloquence quantifications. The point with best trade-off between sensitivity and specificity is indicated as dot. Note that eRV and eRI, being continuous variables, have more steps in ROC curves than EG with three grades.

Figure 4. Functional changes at two months after surgery according to (A) the expected residual tumor volume, (B) the expected resectability index, (C) the eloquence grading, (D) surgical aggressiveness measured as difference between expected and observed residual volume, and (E) surgical aggressiveness measured as difference between extent of resection and resectability index. Each datapoint represents measurements for one patient, in boxes biopsies, and in circles resections. Green datapoints were considered a functional improvement; red datapoints a functional decline.

Figure 5. The diagnostic accuracies for (A) functional decline and (B) functional improvement by the expected residual tumor volume (eRV) in red, the expected resectability index (eRI) in green and the eloquence grade (EG) in blue as receiver operating characteristic curves. The grey curves incorporate age, performance, tumor volume and resection decisions in addition to the eloquence quantifications. The point with best trade-off between sensitivity and specificity is indicated as dot.

Figure 6. Kaplan-Meier survival curves in months after diagnosis for categories of (A) low, intermediate and high expected residual tumor volume, (B) low, intermediate and high expected resectability index, (C) eloquence grade I to III, and (D) biopsies and extents of resection. Censoring is shown for last date of follow-up.

Supplemental Figure 1. Methodology for the expected Residual Volume and expected Resectability Index.

(A, left) For semi-automated volumetric tumor segmentation the standard MRI sequences before and after surgery in patient brain space were imported in Brainlab Smartbrush Suite software (BrainLAB AG, München, Germany). For the tumor segmentations before surgery, we considered gadolinium-enhancing tumor and nonenhancing enclosed necrosis or cyst on T1-weighted images as tumor. The identification of tumor on MRI was guided by the

VASARI MRI feature set, that standardizes radiological review of glioblastoma.(Gevaert, Radiology 2014) For the tumor segmentations after surgery, the aligned T1-weighted images before gadolinium and diffusion-weighted images were reviewed in combination with the post-gadolinium T1-weighted images to distinguish residual disease from surgical artefacts, such as blood and ischemia. For patients who had a biopsy, we defined the residual disease volume as equal to the tumor volume before surgery. The volumetric reconstructions of the segmentations were verified and edited in three orthogonal planes. Initial segmentations were done by one of the authors (DM) and verified by a neurosurgeon (PW) and a neuroradiologist (FB). This step resulted in two sets of binary dicom images for each patient: one set for the tumor volume before surgery and the other for the residual disease volume after surgery in gadolinium-enhanced T1-weighted image space. (A, right) Subsequently, for registration from patient brain space to a common standard brain space, the binary dicom sets were converted to NIfTI format. As standard brain space, we used the Montreal Neurological Institute 1-mm standard brain template based on 152 normal subjects (MNI-152). The pre- and postoperative T1-weighted imaging of each patient were registered towards the MNI brain space using rigid and affine registration, followed by symmetric diffeomorphic registration with B-spline regularization using cross-correlation as similarity metric.(Avants, Medical image analysis 2008) The resulting registered T1-weighted MRIs for each patient were then visually verified and adjusted, if necessary for alignment with the MNI brain template. The registration transformations were subsequently applied to the corresponding binary tumor volumes. This results in a tumor volume before and a residual disease volume after surgery for each patient in standard brain space, such that brain locations are comparable voxel-by-voxel between patients. This processing enables the construction of whole brain maps. (B) For the entire cohort of 915 patients, summation of the number of tumor volumes results in a ‘sum of tumors’ map (green) and summation of the number of residual volumes in a ‘sum of residues’

map. The number of patient volumes contained in a voxel is provided as green-scale and red-scale legends. The resection probability at each voxel is calculated by one minus the division of the sum of residues and the sum of tumors. To account for brain regions that are contralateral and ipsilateral of the tumor, we constructed the probability maps separately for the left and right hemisphere. A resection probability map for the left hemisphere is shown as example and the corresponding legend depicts the probability of resection from 0.0 (red) to 1.0 (green). A value of zero represents a location where tumor was never removed, and a value of one represents a location where tumor was always removed. (C) As an example the calculations at two voxel locations in the brain are highlighted. (D) To calculate the expected residual tumor volume (eRV) and the expected Resectability Index (eRI) for every patient, the resection probability map is masked by the patient's preoperative tumor voxels, here 49,144 tumor voxels representing a volume of 49 mL. The histogram shows the range of resection probabilities at these tumor voxels for this patient. Now the expected resectable volume is calculated by summation of the resection probabilities at these tumor voxels, here representing 35 mL. The preoperative tumor volume minus the expected resectable volume resulted in the eRV, here 14 mL. A division of the expected resectable volume by the preoperative tumor volume resulted in the eRI, here 0.70.

Supplemental Table 1. Patient characteristics by hospital and overall.

Supplemental Table 2. Multivariable logistic regression models of biopsy decisions, resections over 98% and residual volume < 1 mL.

Supplemental Table 3. Multivariable logistic regression models of functional decline and improvement.

Supplemental Table 4. Multivariable proportional hazard models of survival.