

## Quantifying eloquent locations for glioblastoma surgery using resection probability maps

Domenique M. J. Müller, MD,<sup>1</sup> Pierre A. Robe, MD, PhD,<sup>2</sup> Hilko Ardon, MD, PhD,<sup>3</sup> Frederik Barkhof, MD, PhD,<sup>4,5</sup> Lorenzo Bello, MD, PhD,<sup>6</sup> Mitchel S. Berger, MD,<sup>7</sup> Wim Bouwknegt, MD,<sup>8</sup> Wimar A. Van den Brink, MD, PhD,<sup>9</sup> Marco Conti Nibali, MD,<sup>6</sup> Roelant S. Eijgelaar, MSc,<sup>10</sup> Julia Furtner, MD,<sup>11</sup> Seunggu J. Han, MD,<sup>12</sup> Shawn L. Hervey-Jumper, MD,<sup>7</sup> Albert J. S. Idema, MD,<sup>13</sup> Barbara Kiesel, MD,<sup>14</sup> Alfred Kloet, MD,<sup>15</sup> Jan C. De Munck, PhD,<sup>4</sup> Marco Rossi, MD,<sup>6</sup> Tommaso Sciortino, MD,<sup>6</sup> W. Peter Vandertop, MD, PhD,<sup>1</sup> Martin Visser, MSc,<sup>4</sup> Michiel Wagemakers, MD, PhD,<sup>16</sup> Georg Widhalm, MD, PhD,<sup>14</sup> Marnix G. Witte, PhD,<sup>10</sup> Aeilko H. Zwinderman, PhD,<sup>17</sup> and Philip C. De Witt Hamer, MD, PhD<sup>1</sup>

<sup>1</sup>Brain Tumor Center & Department of Neurosurgery, Cancer Center Amsterdam, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands; <sup>2</sup>Department of Neurology & Neurosurgery, University Medical Center Utrecht, The Netherlands; <sup>3</sup>Department of Neurosurgery, St. Elisabeth Hospital, Tilburg, The Netherlands; <sup>4</sup>Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Vrije Universiteit, University Medical Center, Amsterdam, The Netherlands; <sup>5</sup>Institutes of Neurology and Healthcare Engineering, University College London, United Kingdom; <sup>6</sup>Neurosurgical Oncology Unit, Departments of Oncology and Remato-Oncology, Università degli Studi di Milano, Humanitas Research Hospital, IRCCS, Milan, Italy; <sup>7</sup>Department of Neurological Surgery, University of California, San Francisco, California; <sup>8</sup>Department of Neurosurgery, Medical Center Slotervaart, Amsterdam, The Netherlands; <sup>9</sup>Department of Neurosurgery, Isala Hospital, Zwolle, The Netherlands; <sup>10</sup>Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; <sup>11</sup>Department of Biomedical Imaging and Image-Guided Therapy, Medical University Vienna, Austria; <sup>12</sup>Department of Neurological Surgery, Oregon Health and Science University, Portland, Oregon; <sup>13</sup>Department of Neurosurgery, Northwest Clinics, Alkmaar, The Netherlands; <sup>14</sup>Department of Neurosurgery, Medical University Vienna, Austria; <sup>15</sup>Department of Neurosurgery, Medical Center Haaglanden, The Hague, The Netherlands; <sup>16</sup>Department of Neurosurgery, University of Groningen, University Medical Center Groningen, The Netherlands; and <sup>17</sup>Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, The Netherlands

**OBJECTIVE** Decisions in glioblastoma surgery are often guided by presumed eloquence of the tumor location. The authors introduce the “expected residual tumor volume” (eRV) and the “expected resectability index” (eRI) based on previous decisions aggregated in resection probability maps. The diagnostic accuracy of eRV and eRI to predict biopsy decisions, resectability, functional outcome, and survival was determined.

**METHODS** Consecutive patients with first-time glioblastoma surgery in 2012–2013 were included from 12 hospitals. The eRV was calculated from the preoperative MR images of each patient using a resection probability map, and the eRI was derived from the tumor volume. As reference, Sawaya’s tumor location eloquence grades (EGs) were classified. Resectability was measured as observed extent of resection (EOR) and residual volume, and functional outcome as change in Karnofsky Performance Scale score. Receiver operating characteristic curves and multivariable logistic regression were applied.

**RESULTS** Of 915 patients, 674 (74%) underwent a resection with a median EOR of 97%, functional improvement in 71 (8%), functional decline in 78 (9%), and median survival of 12.8 months. The eRI and eRV identified biopsies and EORs of at least 80%, 90%, or 98% better than EG. The eRV and eRI predicted observed residual volumes under 10, 5, and

**ABBREVIATIONS** AUC = area under the curve; EG = eloquence grade; EOR = extent of resection; eRI = expected resectability index; eRV = expected residual tumor volume; KPS = Karnofsky Performance Scale.

**SUBMITTED** November 19, 2019. **ACCEPTED** January 21, 2020.

**INCLUDE WHEN CITING** Published online April 3, 2020; DOI: 10.3171/2020.1.JNS193049.

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

<https://thejns.org/doi/abs/10.3171/2020.1.JNS193049>

**KEYWORDS** glioma; neurosurgery; extent of resection; residual volume; reproducibility of results; oncology

**S**URGICAL 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 hand 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.<sup>1–5</sup> Reports have estimated a minimum threshold for extent of resection (EOR) between 78% and 98%, and a maximum residual volume between 1 and 5 ml to prolong survival of glioblastoma.<sup>6,7</sup> Instead of a threshold, others have reported a continuous positive relationship between resection results and survival.<sup>8,9</sup>

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.<sup>10,11</sup> 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 brain functions noninvasively. Some methods are based on population-based classification of structural anatomy or atlases of brain functions. Sawaya et al.’s classification of eloquence grade (EG) has often been used in reports of surgical cohorts and was shown to correlate with postoperative neurological deficits.<sup>12</sup> Other surgical classifications exist for the whole brain<sup>1,13</sup> or brain regions.<sup>14</sup> Brain atlases have been constructed for brain plasticity<sup>15</sup> and cortical<sup>16</sup> and subcortical<sup>17</sup> structures. Other methods are based on localizing brain function in individuals, such as preoperative diagnostic imaging with structural MRI,<sup>18</sup> task-based functional MRI,<sup>19</sup> resting-state functional MRI,<sup>20</sup> magnetoencephalography,<sup>21</sup> white matter tractography,<sup>22</sup> or transcranial magnetic stimulation.<sup>23</sup>

Another source of potentially useful information before surgery is resection probability maps based on a large number of prior resections. Resection probability maps of nonenhancing glioma have been previously used to estimate the resection result,<sup>3,24</sup> to evaluate the potential for brain plasticity,<sup>15</sup> and to compare resection results between surgical teams.<sup>25,26</sup> A new application for resection prob-

ability 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, estimate resectability, and predict functional outcome and survival in comparison with EG.

## Methods

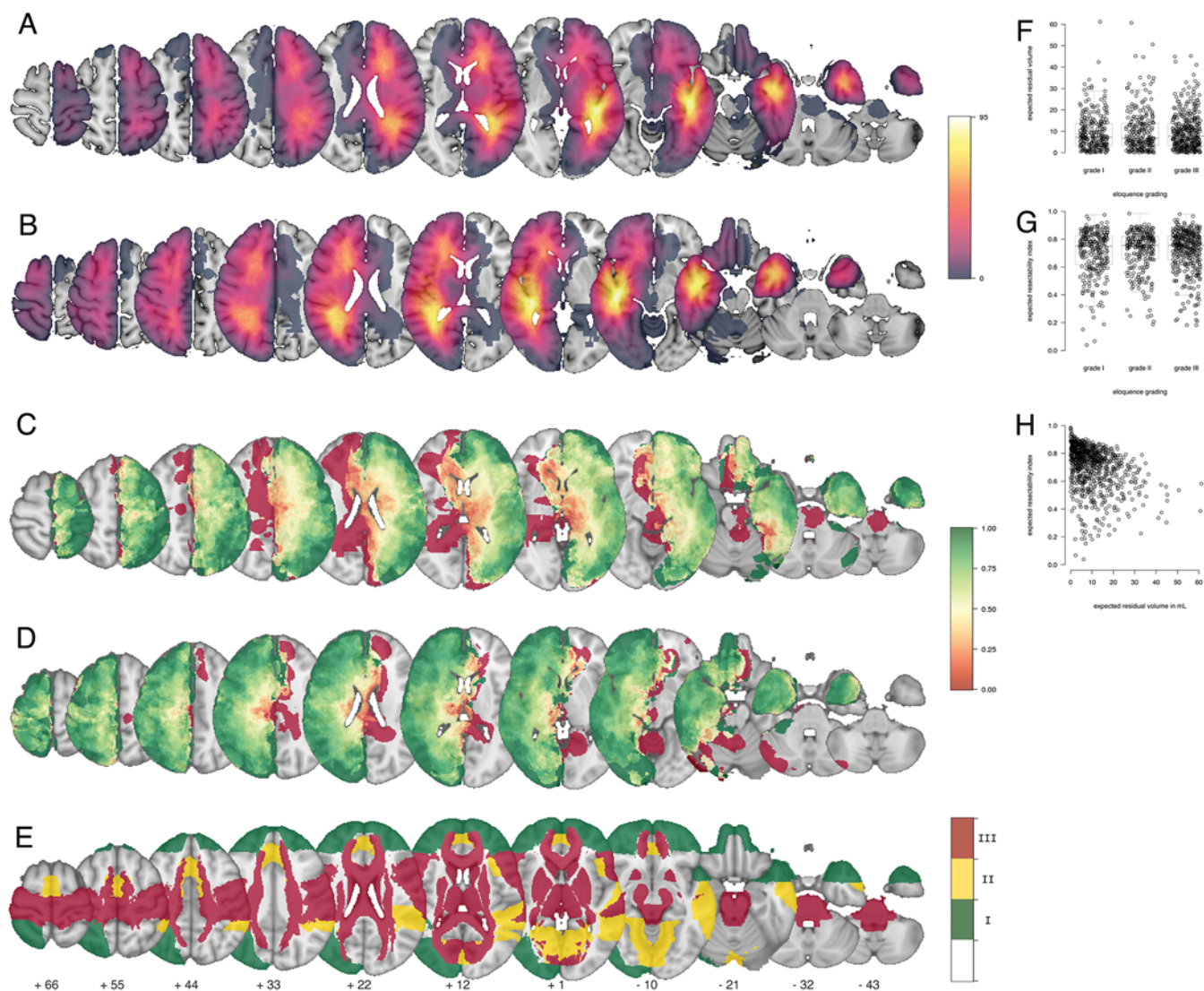
### Patients

We identified all patients at least 18 years of age with a newly diagnosed glioblastoma at first-time surgery between January 1, 2012, and December 31, 2013, from 12 hospitals: North West Hospital, Alkmaar, The Netherlands; Slotervaart Hospital, Amsterdam, The Netherlands; Isala Hospital, Zwolle, The Netherlands; Humanitas Research Hospital Milano, Italy; University Medical Center Groningen, The Netherlands; Medical University Vienna, Austria; Hôpital Lariboisière, Paris, France; Medical Center Haaglanden, The Hague, The Netherlands; Amsterdam University Medical Centers, The Netherlands; University of California, San Francisco Medical Center, San Francisco, California; St. Elisabeth Hospital, Tilburg, The Netherlands; and University Medical Center Utrecht, The Netherlands. 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 nonenhancing glioblastoma on MRI, if MRI was performed later than 3 days postoperatively,<sup>27</sup> or if the imaging was of insufficient quality for segmentation and registration due to missing images or severe motion artifacts. 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, sex, type of surgery, the Karnofsky Performance Scale (KPS) score preoperatively and at 2 months postoperatively, and the dates of surgery, last follow-up in the clinic, and death. We also obtained the MR images pre- and postoperatively.

Imaging protocols were standardized in hospitals, although not identical between hospitals, and included: 1) 3D heavily T1-weighted gradient-echo pulse sequence at 1 mm isotropic resolution (at most 5 mm), obtained after administration of intravenous gadolinium; 2) a T2/FLAIR-weighted gradient-echo pulse sequence; and 3) a diffusion-weighted sequence.

Approval of the study protocol was obtained by the IRBs and informed consent from patients was obtained



**FIG. 1.** **A and B:** The tumor probability maps for glioblastoma in the left (A,  $n = 451$ ) and right (B,  $n = 464$ ) hemisphere, indicating the number of cases in the data set involved by tumor within each voxel. **C and D:** Resection probability maps for the left (C) and right (D) hemisphere indicating the resectability index at 1-mm resolution. **E:** The EG in standard brain space. The numbers across the bottom refer to Z-values of standard brain space. **F–H:** The correlation per patient between the EG and the eRV (F) and the eRI (G), and between the eRV and the eRI (H). The EG regions were specified by Sawaya et al.<sup>12</sup> The color codes correspond to the legends: in panels A and B, 95 is the maximum number of cases per voxel. Figure is available in color online only.

according to local regulations. Data and imaging for analysis were anonymized.

### Expected Residual Tumor Volume, Expected Resectability Index, and Eloquent Grade

Surgical eloquence was quantitated using a resection probability map derived from all patients. For details on the methodology see Supplementary Fig. 1 and our earlier description.<sup>26</sup> This resection probability map represents the probability of resection at 1 mm<sup>3</sup> resolution in standard brain space, i.e., the MNI-152 brain template,<sup>28</sup> based on pre- and postoperative tumor segmentations.<sup>24–26</sup> We used software (Smartbrush Suite Software; Brainlab AG) to outline the contrast-enhancing elements on the T1-

weighted 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 artifacts from residual tumor in patients who had undergone 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 resid-

ual tumor volume (eRV), expressed in milliliters.<sup>3</sup> 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 nonresectable tumor and a value of 1.0 a completely resectable tumor. A web application was created to enable calculations of the eRV and eRI from preoperative MR images of new patients (<https://www.pictureproject.nl>).

In addition, we classified the EG for each patient's tumor, ranging from noneloquent (grade I), to near-eloquent (grade II), to eloquent (grade III) as a historical reference.<sup>12</sup> 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 approach we determined for each patient the number of tumor voxels overlapping with the regions corresponding to these eloquence grade definitions,<sup>12</sup> as indicated in Fig. 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 EOR, 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, sex, preoperative KPS score, and tumor volume. For the observed EORs, we calculated the postoperative residual volume and preoperative tumor volume. The percentage of the preoperative volume that had been resected was considered the EOR. We evaluated the eRV, eRI, and EG in their prediction of a range of EORs of more than 80%, 90%, and 98%, and a range of observed residual tumor volumes of less than 10, 5, and 1 ml. For the postoperative KPS score changes, we subtracted the performance at 2 months postoperative from the preoperative performance. An increase in KPS score of 20 points or more was considered a functional improvement; a decrease in KPS score of 20 points or less was considered a functional decline.<sup>29</sup> We related postoperative performance change with the eRV, eRI, and EG together with age, sex, preoperative performance, and tumor volume. To determine 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 sex 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 characteristic plots. We calculated the areas under the curve (AUCs) of sensitivity versus specificity and compared AUCs using the bootstrap method;<sup>30</sup> these were interpreted as poor, rea-

sonable, good, and excellent.<sup>31</sup> The *p* values < 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 a binary response variable and eRV, eRI, EG, age, sex, preoperative performance, and tumor volume as dependent variables. The regression assumptions were visually verified and met. We used survival analysis with survival time as a response variable; eRV, eRI, and EG as predictors; and age, sex, preoperative performance, and tumor volume as confounders in a 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

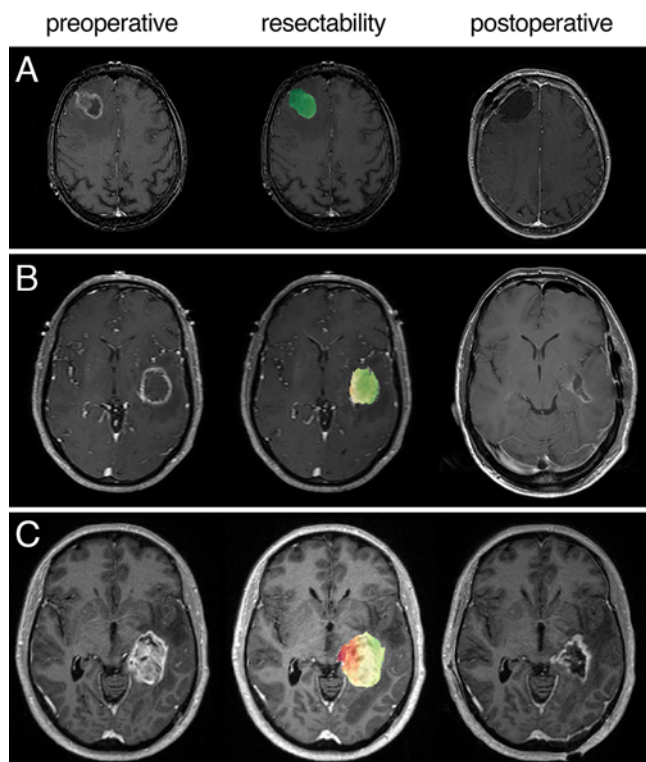
A total of 1083 patients were identified and evaluated for inclusion. Of these, 110 patients were excluded because the MR image was later than 3 days postoperatively; 53 patients had MR images of insufficient quality to register to standard space; and 5 patients had tumors that did not enhance with contrast administration. The characteristics of the 915 patients for analysis are listed in Supplementary Table 1. Patient characteristics were comparable across centers.

The tumor probability map, the resection probability map, and the EG are shown in Fig. 1. The EG did not correlate with the eRV (Kendall's tau = 0.01, Fig. 1F) or with the eRI (Kendall's tau = 0.02, Fig. 1G). The eRV and eRI were correlated (Kendall's tau = -0.28, *p* < 0.00001, Fig. 1H). Three examples of decreasing resectability are demonstrated in Fig. 2.

### Biopsy Indications and Resectability

Overall, 241 patients (26%) were biopsied and 674 patients (74%) underwent a resection, with a median EOR of 97% (Supplementary Table 1).

To evaluate the relation among eRV, eRI, and EG to predict resectability, we first determined the diagnostic accuracy to identify patients who underwent biopsy (Fig. 3A). The eRI identified biopsies significantly better (AUC 0.77, 95% CI 0.74–0.81, *p* < 0.00001) than the eRV (AUC 0.52, 95% CI 0.48–0.56) or the EG (AUC 0.47, 95% CI 0.43–0.51). Although the discrimination using eRI is good, a positive predictive value of 48% was observed, 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 EOR of at least 98%, 90%, and 80% for patients with a resection (Fig. 3B). The eRI was significantly better at identifying resections of at least 98% (AUC 0.61, 95% CI 0.57–0.65, *p* < 0.00001), 90% (AUC 0.65, 95% CI 0.61–0.69, *p* < 0.00001), and 80% (AUC 0.66, 95% CI 0.60–0.71, *p* = 0.00005) as was the eRV (AUC 0.60, 95% CI 0.55–0.63, *p* = 0.0003; AUC 0.58, 95% CI 0.53–0.63, *p* = 0.005; and AUC 0.56, 95% CI 0.50–0.62, *p* value nonsignificant, respectively) compared with the EG (AUC 0.49, 95% CI



**FIG. 2.** Axial MR images showing case examples of decreasing resectability. **A:** A 74-year-old man with a 13-ml tumor, an eRV of 1.2 ml, an eRI of 0.92, an EG of I, an observed residual volume of 0 ml, and 100% EOR. **B:** A 63-year-old woman with a 49-ml tumor, an eRV of 14 ml, an eRI of 0.70, an EG of II, an observed residual volume of 2.6 ml, and an EOR of 95%. **C:** A 47-year-old man with a 40-ml tumor, an eRV of 22 ml, an 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 Fig. 1. Figure is available in color online only.

0.45–0.53; AUC 0.50, 95% CI 0.45–0.54; and AUC 0.48, 95% CI 0.43–0.54, respectively). Again, the discrimination using the eRI to identify patients with an EOR of 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 of predicting residual volume of less than 10, 5, and 1 ml for patients with a resection (Fig. 3C). The eRV was significantly better at identifying observed residual volumes less than 10 ml (AUC 0.81, 95% CI 0.77–0.85,  $p < 0.00001$ ), 5 ml (AUC 0.77, 95% CI 0.73–0.81,  $p < 0.00001$ ), and 1 ml (AUC 0.73, 95% CI 0.69–0.76,  $p < 0.00001$ ) as was the eRI (AUC 0.66, 95% CI 0.60–0.72,  $p = 0.003$ ; AUC 0.65, 95% CI 0.61–0.70,  $p < 0.00001$ ; and AUC 0.63, 95% CI 0.59–0.67,  $p < 0.00001$ , respectively) compared with the EG (AUC 0.53, 95% CI 0.46–0.59; AUC 0.49, 95% CI 0.44–0.54; and AUC 0.49, 95% CI 0.45–0.54, respectively). The discrimination using the 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 for identifying

the biopsy indications (Fig. 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 (Supplementary 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 (Fig. 3A; AUC 0.84, 95% CI 0.82–0.87,  $p = 0.0022$ ).

The eRV and eRI are also imperfect at identifying the patients with resections > 98% (Fig. 3B). Lower eRV, higher eRI, and decreasing EG had a strong association with resections > 98% in conjunction with age, performance, and tumor volume (Supplementary Table 2). Resections > 98% based on the eRI were significantly improved by combining the clinical variables (Fig. 3B; AUC 0.68, 95% CI 0.64–0.72,  $p = 0.019$ ). Furthermore, the eRV was better than the eRI in identifying resections with less than 1 ml residual volume, which could not be improved further with the clinical variables (Fig. 3C; Supplementary Table 2).

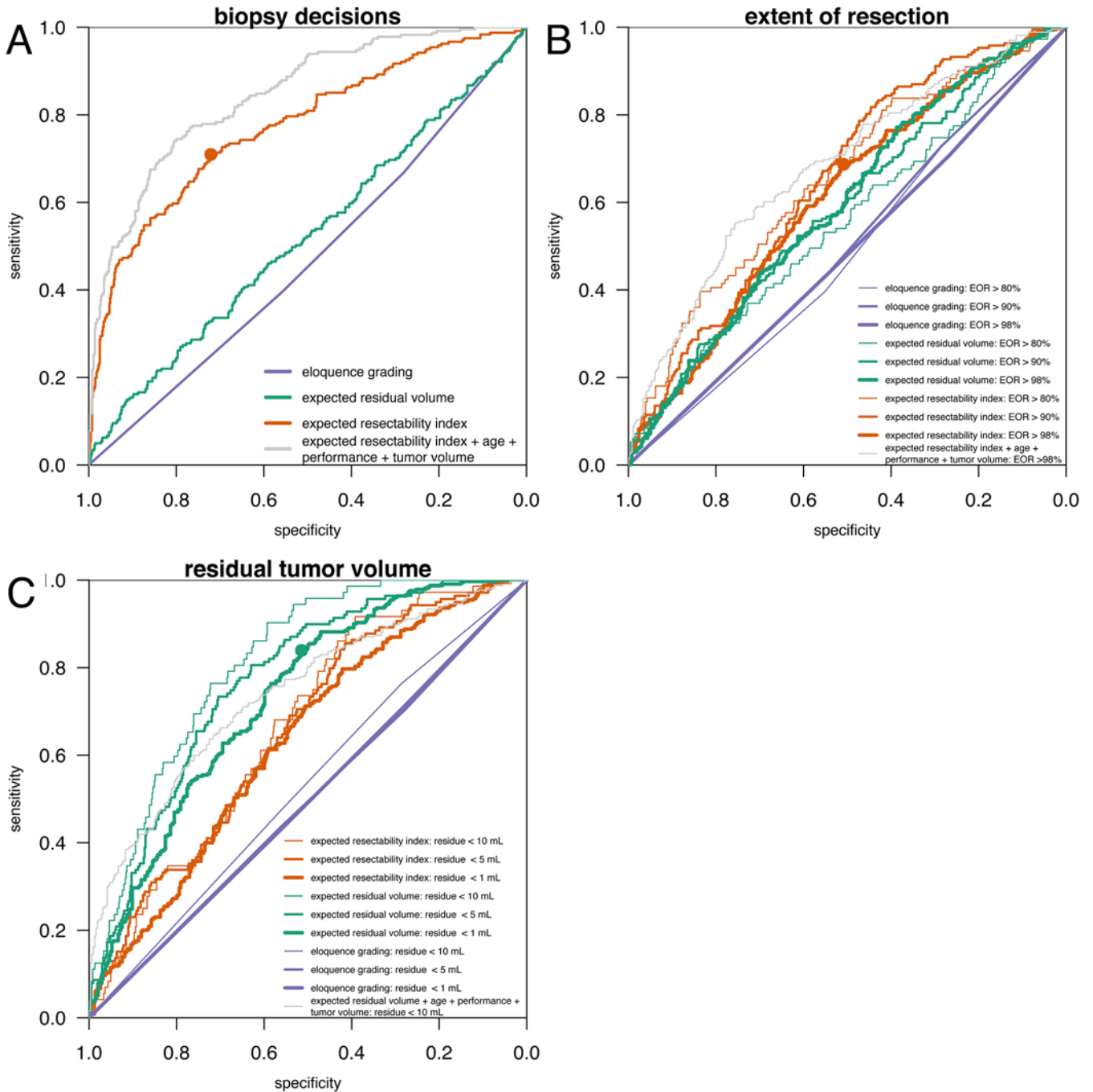
### Functional Outcome

Overall, 71 patients (8%) improved functionally and 78 (9%) declined between surgery and 2 months follow-up. Functional change information was missing in 90 patients (10%).

The eRV, eRI, and EG were not related according to 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 (Fig. 4A–C). An increased odds of functional decline was significantly related to a lower eRI in logistic regression, as well as with higher preoperative KPS score and higher tumor volume (Supplementary 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 (Supplementary 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 eRI from the observed EOR. These measures of surgical aggressiveness clearly separate biopsies from resections (Fig. 4D and E), but do not indicate that surgical aggressiveness is associated with functional changes.

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



**FIG. 3.** The diagnostic accuracies for biopsy decisions (A), EOR (B), and observed residual tumor volume (C) by the eRV in green, the eRI in red, and the EG in purple as receiver operating characteristic curves. The gray curves incorporate age, performance, and tumor volume in addition to the eloquence quantifications. The point with the best trade-off between sensitivity and specificity is indicated by a dot. Note that eRV and eRI, being continuous variables, have more steps in receiver operating characteristic curves than EG with three grades. Figure is available in color online only.

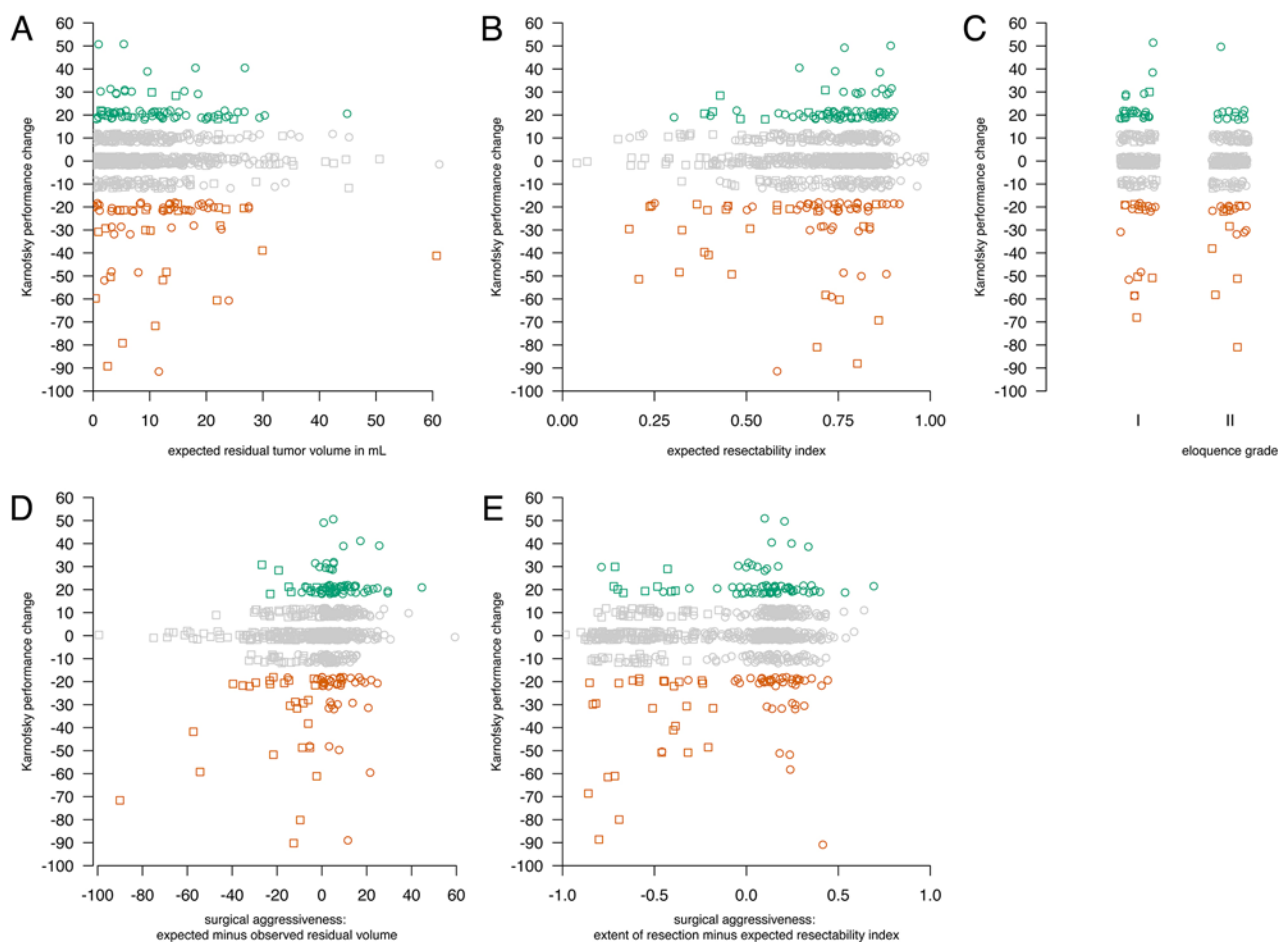
similarly poor (AUC 0.57, 95% CI 0.50–0.63; AUC 0.58, 95% CI 0.51–0.64; and AUC 0.51, 95% CI 0.44–0.58, respectively).

Adding the available clinical information of age, performance, tumor volume, and resection decision significantly improved the identification of functional improvement (AUC 0.89, 95% CI 0.85–0.92,  $p < 0.00001$ ; Fig. 5B)

and to a lesser extent the identification of functional decline (AUC 0.70, 95% CI 0.64–0.76,  $p = 0.036$ ; Fig. 5A).

### Survival

The median overall survival was 12.8 months for the whole population. To evaluate eRV, eRI, and EG and their association with survival, we first plotted Kaplan-Meier



**FIG. 4.** Functional changes measured by the change in KPS score at 2 months after surgery compared to baseline according to the eRV (A), the eRI (B), and the EG (C); surgical aggressiveness measured as the difference between expected and observed residual volume (D); and surgical aggressiveness measured as the difference between the EOR and the eRI (E). Each *data point* represents measurements for 1 patient; biopsies are represented by *boxes*, and resections by *circles*. *Green data points* were considered a functional improvement, whereas *red data points* were considered a functional decline. Figure is available in color online only.

curves (Fig. 6). The eRV was significantly associated with survival (Fig. 6A,  $p < 0.00001$ ) with median overall survival times of 16, 13, and 9 months for low, intermediate, and high eRV, respectively, at cutoffs of 1.7 and 17 ml. The eRI was also significantly associated with survival (Fig. 6B,  $p < 0.00001$ ) with median overall survival times of 16, 13, and 8 months for high, intermediate, and low eRI, respectively, at cutoffs of 0.8 and 0.5. The EG was not associated with survival (Fig. 6C,  $p$  value nonsignificant). For comparison, actual biopsy decision and observed EOR—only available after surgery—were confirmed to be strongly associated with survival (Fig. 6D,  $p < 0.00001$ ) with median overall survival times of 18, 13, and 6 months for EORs  $> 98\%$ ,  $< 98\%$ , and biopsies.

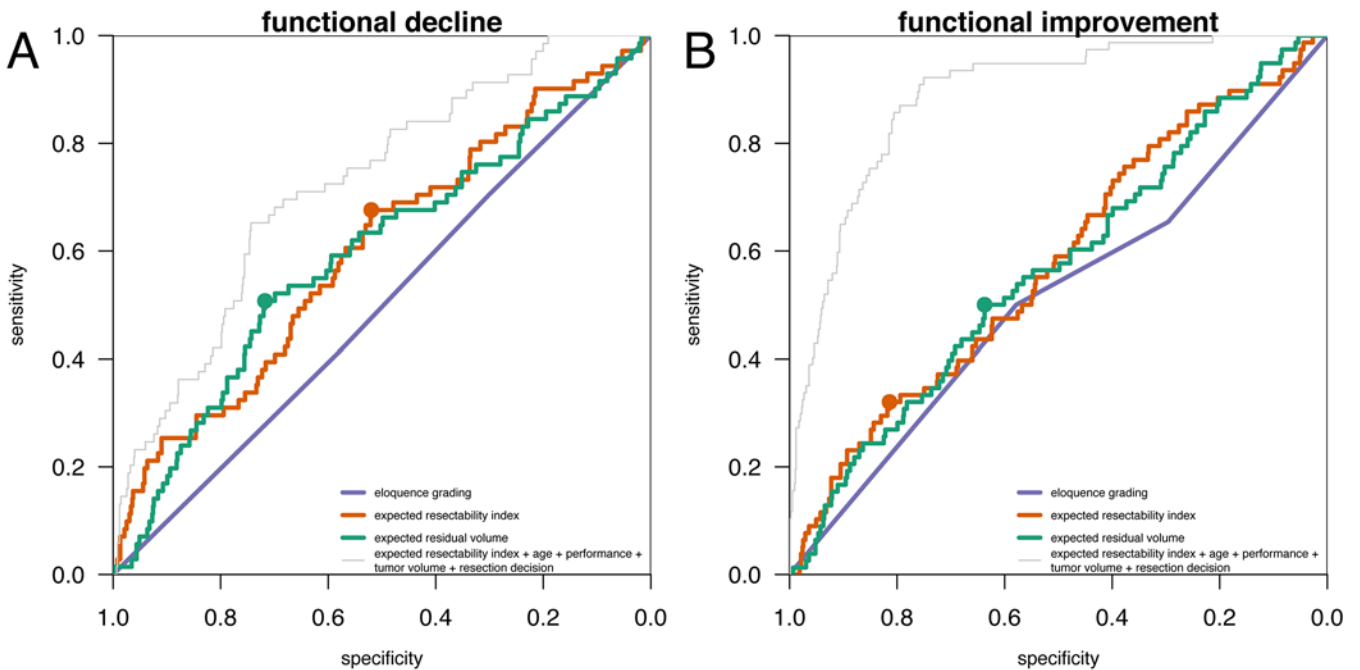
To determine the interaction between the eRV and eRI and known prognostic factors, we further explored this association in multivariable models (Supplementary Table 4). A larger eRV and a lower eRI were significantly associated with shorter survival, independent from greater age, lower performance, and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

firmed 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 EOR. 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



**FIG. 5.** The diagnostic accuracies for functional decline (**A**) and functional improvement (**B**) by the eRV in *green*, the eRI in *red*, and the EG in *purple* as receiver operating characteristic curves. The *gray* curves incorporate age, performance, tumor volume, and resection decisions in addition to the eloquence quantifications. The point with the best trade-off between sensitivity and specificity is indicated by a *dot*. Figure is available in color online only.

critical regions of the brain. Resection probability maps can be used in combination with other methods, such as brain atlases of positive stimulation sites,<sup>17</sup> 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 identification using electrostimulation during surgery. Resectability, functional outcome, and survival have been related to the positive stimulation areas<sup>2,10</sup> as well as with eloquence grading.<sup>6,7,32,33</sup> 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.

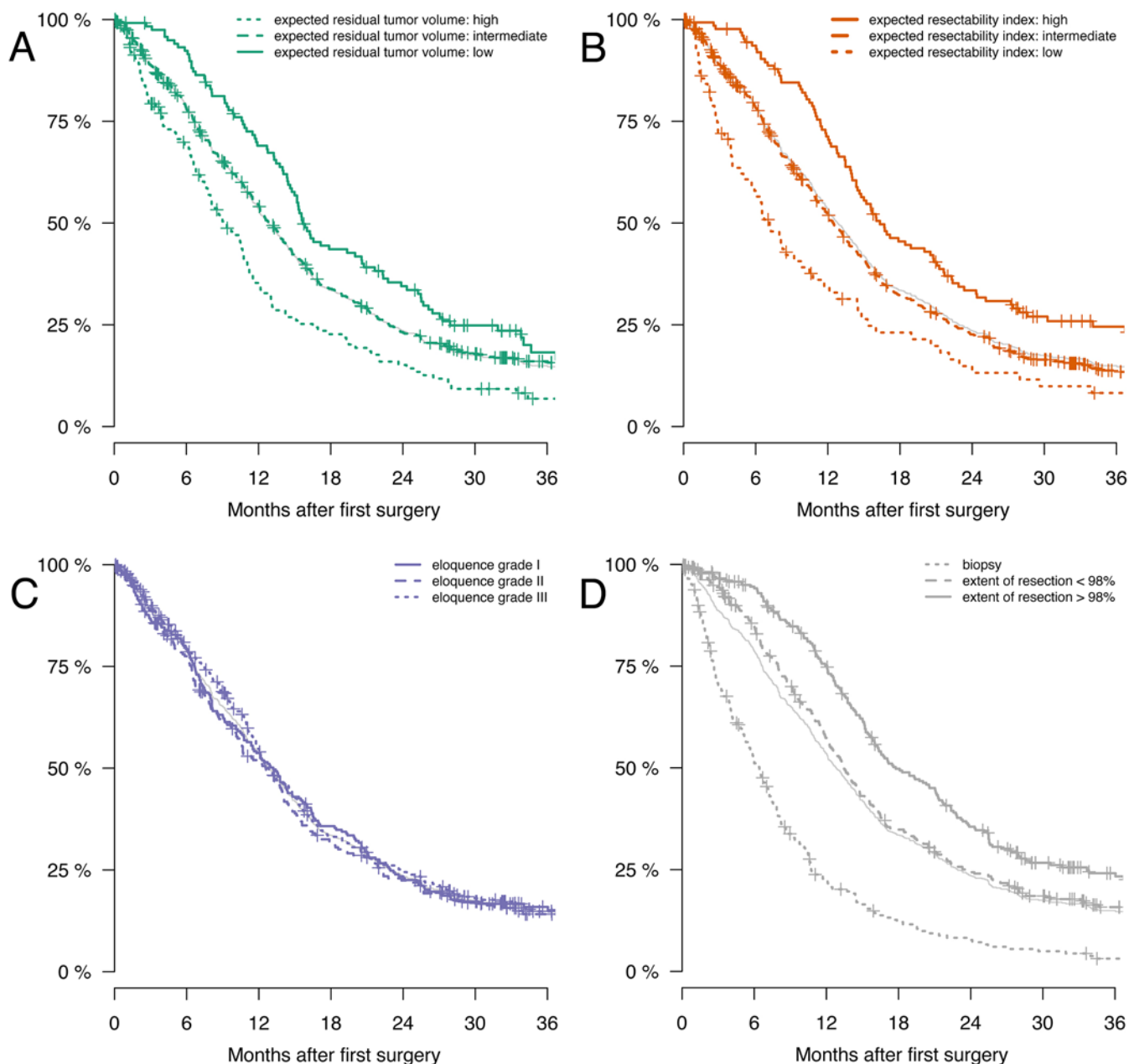
The eRV and eRI are probably most useful for clinical decisions before surgery. Surgical decisions are now based on intuitive estimates of resectability with input from presumed functionality of tumor-infiltrated brain regions, with associated risk for functional deficits and from a presumed increase in life expectancy by removing more tumor tissue.<sup>9</sup> The eRV and eRI appear to improve the EG, which, 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. Notwithstand-

ing the less than optimal accuracy, these predictions may improve current practice, given the large variation in rates of biopsies and gross-total resections among reports of large populations,<sup>34–36</sup> which suggests considerable treatment 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.<sup>4,9,12</sup> 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 guidance<sup>22</sup> and intraoperative stimulation mapping.<sup>10</sup> For instance, several neurosurgical teams have replaced biopsies with resections using 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.<sup>37</sup> 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





**FIG. 6.** Kaplan-Meier survival curves in months after first surgery for categories of low, intermediate, and high eRV (A); low, intermediate, and high eRI (B); EG I–III (C); and biopsies and EORs (D). Censoring is shown for the last date of follow-up. Figure is available in color online only.

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 the difference between the expected and observed tumor removal—is not associated with functional changes, supports the idea that neurosurgeons successfully 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 conse-

quently, 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.<sup>38,39</sup> 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 zones<sup>40</sup> and white matter intersections<sup>41</sup> have been associated with poor survival despite maximal resections.

The strengths of this study are that well-documented complete patient populations from 12 neurooncological 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 standards<sup>42</sup> by segmenting the gadolinium-enhancing portions of glioblastoma, which is known to undersample the true extent of glioblastoma.<sup>43</sup> In addition, neuroimaging protocols, manual tumor segmentation,<sup>44</sup> and brain registration<sup>45</sup> can be subject to variation and are therefore potential sources of error, although visual inspection verified satisfactory processing in all patients. The KPS score 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.

## Acknowledgments

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 conducted 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. This research is part of the Innovative Medical Devices Initiative program with project number 10-10400-96-14003, which is financed by the Netherlands Organisation for Scientific Research (NWO). This research is also supported by a research grant from the Dutch Cancer Society (no. VU2014-7113). Furthermore, this research is supported by the National Institute for Health Research of the University College London Hospitals (UCLH) Biomedical Research Centre.

## References

1. Capellades J, Puig J, Domenech S, 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.* 2018;137(2):367–377.
2. Coburger J, Renovanz M, Ganslandt O, et al. Evaluation of surgical decision making and resulting outcome in patients with highly eloquent glioblastoma: results of a multicenter assessment. *Clin Neurol Neurosurg.* 2017;162:29–35.
3. Hendriks EJ, Habets EJJ, Taphoorn MJB, et al. Linking late cognitive outcome with glioma surgery location using resection cavity maps. *Hum Brain Mapp.* 2018;39(5):2064–2074.
4. Sonabend AM, Zacharia BE, Cloney MB, et al. Defining glioblastoma resectability through the wisdom of the crowd: a proof-of-principle study. *Neurosurgery.* 2017;80(4):590–601.
5. Southwell DG, Birk HS, Han SJ, et al. Resection of gliomas deemed inoperable by neurosurgeons based on preoperative imaging studies. *J Neurosurg.* 2018;129(3):567–575.
6. Chaichana KL, Jusue-Torres I, Navarro-Ramirez R, et al. Establishing percent resection and residual volume thresholds affecting survival and recurrence for patients with newly diagnosed intracranial glioblastoma. *Neuro Oncol.* 2014;16(1):113–122.
7. Sanai N, Polley M-Y, McDermott MW, et al. An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg.* 2011;115(1):3–8.
8. Ellingson BM, Abrey LE, Garcia J, et al. Post-chemoradiation volumetric response predicts survival in newly diagnosed glioblastoma treated with radiation, temozolomide, and bevacizumab or placebo. *Neuro Oncol.* 2018;20(11):1525–1535.
9. Marko NF, Weil RJ, Schroeder JL, et al. 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.* 2014;32(8):774–782.
10. De Witt Hamer PC, Robles SG, Zwinderman AH, et al. Impact of intraoperative stimulation brain mapping on glioma surgery outcome: a meta-analysis. *J Clin Oncol.* 2012;30(20):2559–2565.
11. Duffau H. Stimulation mapping of white matter tracts to study brain functional connectivity. *Nat Rev Neurol.* 2015;11(5):255–265.
12. Sawaya R, Hammoud M, Schoppa D, et al. Neurosurgical outcomes in a modern series of 400 craniotomies for treatment of parenchymal tumors. *Neurosurgery.* 1998;42(5):1044–1056.
13. Shinoda J, Sakai N, Murase S, et al. Selection of eligible patients with supratentorial glioblastoma multiforme for gross total resection. *J Neurooncol.* 2001;52(2):161–171.
14. Hervey-Jumper SL, Li J, Osorio JA, 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.* 2016;124(2):482–488.
15. Herbet G, Maheu M, Costi E, et al. Mapping neuroplastic potential in brain-damaged patients. *Brain.* 2016;139(Pt 3):829–844.
16. Tate MC, Herbet G, Moritz-Gasser S, et al. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain.* 2014;137(Pt 10):2773–2782.
17. Sarubbo S, De Benedictis A, Merler S, et al. Towards a functional atlas of human white matter. *Hum Brain Mapp.* 2015;36(8):3117–3136.
18. Talos I-F, Zou KH, Ohno-Machado L, et al. Supratentorial low-grade glioma resectability: statistical predictive analysis based on anatomic MR features and tumor characteristics. *Radiology.* 2006;239(2):506–513.
19. Castellano A, Cirillo S, Bello L, et al. Functional MRI for surgery of gliomas. *Curr Treat Options Neurol.* 2017;19(10):34.

20. Cochereau J, Deverdun J, Herbet G, et al. Comparison between resting state fMRI networks and responsive cortical stimulations in glioma patients. *Hum Brain Mapp*. 2016;37(11):3721–3732.
21. Baillet S. Magnetoencephalography for brain electrophysiology and imaging. *Nat Neurosci*. 2017;20(3):327–339.
22. Barone DG, Lawrie TA, Hart MG. Image guided surgery for the resection of brain tumours. *Cochrane Database Syst Rev*. 2014;1(1):CD009685.
23. Krieg SM, Lioumis P, Mäkelä JP, et al. Protocol for motor and language mapping by navigated TMS in patients and healthy volunteers; workshop report. *Acta Neurochir (Wien)*. 2017;159(7):1187–1195.
24. Mandonnet E, Jbabdi S, Taillandier L, et al. Preoperative estimation of residual volume for WHO grade II glioma resected with intraoperative functional mapping. *Neuro Oncol*. 2007;9(1):63–69.
25. De Witt Hamer PC, Hendriks EJ, Mandonnet E, et al. Resection probability maps for quality assessment of glioma surgery without brain location bias. *PLoS One*. 2013;8(9):e73353.
26. Müller DMJ, Robe PAJT, Eijgelaar RS, et al. Comparing glioblastoma surgery decisions between teams using brain maps of tumor locations, biopsies, and resections. *JCO Clin Cancer Inform*. 2019;3:1–12.
27. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol*. 2010;28(11):1963–1972.
28. Fonov VS, Evans AC, McKinstry RC, et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*. 2009;47(Suppl 1):S102.
29. Chaichana KL, Halthore AN, Parker SL, et al. Factors involved in maintaining prolonged functional independence following supratentorial glioblastoma resection. Clinical article. *J Neurosurg*. 2011;114(3):604–612.
30. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
31. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. *EJIFCC*. 2009;19(4):203–211.
32. 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*. 2016;124(4):977–988.
33. Smith JS, Chang EF, Lamborn KR, et al. Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas. *J Clin Oncol*. 2008;26(8):1338–1345.
34. Bauchet L, Mathieu-Daudé H, Fabbro-Peray P, et al. Oncological patterns of care and outcome for 952 patients with newly diagnosed glioblastoma in 2004. *Neuro Oncol*. 2010;12(7):725–735.
35. Graus F, Bruna J, Pardo J, et al. Patterns of care and outcome for patients with glioblastoma diagnosed during 2008–2010 in Spain. *Neuro Oncol*. 2013;15(6):797–805.
36. Scoccianti S, Magrini SM, Ricardi U, 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*. 2010;67(2):446–458.
37. Gempt J, Förschler A, Buchmann N, et al. Postoperative ischemic changes following resection of newly diagnosed and recurrent gliomas and their clinical relevance. *J Neurosurg*. 2013;118(4):801–808.
38. Ellingson BM, Lai A, Harris RJ, et al. Probabilistic radiographic atlas of glioblastoma phenotypes. *AJNR Am J Neuroradiol*. 2013;34(3):533–540.
39. Tejada Neyra MA, Neuberger U, Reinhardt A, et al. Voxel-wise radiogenomic mapping of tumor location with key molecular alterations in patients with glioma. *Neuro Oncol*. 2018;20(11):1517–1524.
40. Chaichana KL, McGirt MJ, Frazier J, et al. Relationship of glioblastoma multiforme to the lateral ventricles predicts survival following tumor resection. *J Neurooncol*. 2008;89(2):219–224.
41. Mickevicius NJ, Carle AB, Bluemel T, et al. Location of brain tumor intersecting white matter tracts predicts patient prognosis. *J Neurooncol*. 2015;125(2):393–400.
42. Ellingson BM, Wen PY, Cloughesy TF. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics*. 2017;14(2):307–320.
43. Jain R, Poisson LM, Gutman D, et al. Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. *Radiology*. 2014;272(2):484–493.
44. Visser M, Müller DMJ, van Duijn RJM, et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *Neuroimage Clin*. 2019;22:101727.
45. Ou Y, Akbari H, Bilello M, et al. Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights. *IEEE Trans Med Imaging*. 2014;33(10):2039–2065.

## Disclosures

Brainlab provided segmentation software as a contribution in kind to this study.

## Author Contributions

Conception and design: De Witt Hamer, Müller, Robe, Barkhof, Berger, Vandertop. Acquisition of data: De Witt Hamer, Müller, Eijgelaar, Furtner, De Munck, Visser, Witte, Zwinderman. Analysis and interpretation of data: De Witt Hamer, Müller, Eijgelaar, De Munck, Visser, Witte, Zwinderman. Drafting the article: De Witt Hamer, Müller. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: De Witt Hamer. Statistical analysis: De Witt Hamer, Müller, Zwinderman. Administrative/technical/material support: Müller. Study supervision: De Witt Hamer.

## Supplemental Information

### Online-Only Content

Supplemental material is available with the online version of the article.

*Supplementary Figure and Tables.* <https://thejns.org/doi/suppl/10.3171/2020.1.JNS193049>.

## Correspondence

Philip C. De Witt Hamer: Amsterdam University Medical Centers, Amsterdam, The Netherlands. [p.dewitthamer@vumc.nl](mailto:p.dewitthamer@vumc.nl)