Preoperative resectability estimates of non-enhancing glioma by neurosurgeons and a resection probability map

Eef J. Hendriks MD [1], Sander Idema MD, PhD [2], Shawn L. Hervey-Jumper MD [3], Anne-Laure Bernat MD [4], Aeilko H. Zwinderman PhD [5], Frederik Barkhof MD, PhD [1,6], W. Peter Vandertop MD, PhD [2], Emmanuel Mandonnet MD, PhD [4], Hugues Duffau MD, PhD [7], Mitchel S. Berger MD [3], Philip C. De Witt Hamer MD, PhD [2]

version: July 24, 2018

Affiliations

1 Department of Radiology & Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands; 2 Neurosurgical Center Amsterdam, VU University Medical Center and Academic Medical Center, Amsterdam, The Netherlands; 3 Department of Neurological Surgery, University of California, San Francisco, San Francisco, California; 4 Department of Neurosurgery, Hôpital de la Pitié-Salpêtrière, Paris, France; 5 Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, Netherlands; 6 Institutes of Neurology & Healthcare Engineering, UCL, London, UK; 7 Department of Neurosurgery, Hôpital Gui de Chauliac, Centre Hospitalier Universitaire Montpellier, Montpellier, France.

Correspondence

Dr Philip C. De Witt Hamer VU University Medical Center, Dept of Neurosurgery De Boelelaan 1117, 1081 HV Amsterdam, Netherlands Contact T: +31 20 444 3783, F: +31 20 444 3784, E: p.dewitthamer@vumc.nl

Funding

This research is part of the programme Innovative Medical Devices Initiative 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 (VU2014-7113). Furthermore, this research is supported by the National Institute for Health Research of the University College London Hospitals (UCLH) Biomedical Research Centre.

Conflict of Interest

No conflicts of interest.

Acknowledgements

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).

Presentation

Plenary presentation on the Annual Winter Meeting of the Dutch Society for Neurosurgery, 26th of January 2018, Utrecht, The Netherlands.

ABSTRACT

Background. Preoperative interpretation of resectability of diffuse nonenhancing glioma is primarily based on individual surgical expertise.

Objective. To compare the accuracy and precision between observed resections and preoperative estimates of neurosurgeons and a resection probability map (RPM). We hypothesize that the RPM estimates is as good as senior neurosurgeons.

Methods. A total of 234 consecutive patients were included from two centers, who had resective surgery with functional mapping between 2006 and 2012 for a supra-tentorial nonenhancing glioma. Extent of resection (EOR) and residual tumor volume (RTV) were segmented and an RPM was constructed in standard brain space. Three junior and three senior neurosurgeons estimated EOR and RTV, blinded for postoperative results. We determined the agreement between the estimates and calculated the diagnostic accuracy of the neurosurgeons and the RPM to predict observed resections.

Results. Preoperative estimates of resection results by junior and senior neurosurgeons were significantly biased towards overestimation of EOR (4.2% and 11.2%) and underestimation of RTV (4.3 and 9.0 mL), whereas estimates of the RPM were unbiased (-2.6% and -0.2 mL, respectively). The limits of agreement were wide for neurosurgeons and for the RPM. The RPM was significantly more accurate in identifying patients in whom an EOR>40% was observed than neurosurgeons.

Conclusion. Neurosurgeons estimate preoperative resectability before surgery of a nonenhancing glioma rather accurate - with a small bias - and imprecise - with wide limits of agreement. An RPM provides unbiased resectability estimates, which can be useful for surgical decision- making, planning, and education.

Running title: Preoperative resectability of nonenhancing glioma

Keywords: glioma; neurosurgery; extent of resection; residual tumor volume; resection probability map

INTRODUCTION

Nonenhancing glioma is often of WHO grade II, sometimes of grade III,[1] for which surgery, radiotherapy, and chemotherapy are treatment options[2,3]. More extensive surgical removal is associated with longer survival.[4–8]

Surgical decision-making aims to identify those patients who should have more benefit than risk from resection. Moreover, it is important to avoid unnecessary surgical risk by preselecting those patients for biopsy in whom a reasonable resection of extensive disease altering the outcome is not possible. This decision is based on a preoperative interpretation of resectability, based on the patient's age and condition, and the tumor's location and volume. A minimum extent of resection (EOR) or residual volume to improve the natural course of nonenhancing glioma remains undetermined, in contrast to reports on glioblastoma.[5,9,10] Nevertheless, several reports suggest that patients with partial resection of a low-grade glioma of at least 40%, 70%, or 90% EOR or at most 10, 15, or 20 mL of residual tumor volume live longer [4,11–13] with fewer seizures.[14]

The preoperative interpretation of resectability is primarily based on individual surgical expertise, with the potential of subjectiveness, supported by the discrepancy between the individual surgeon's impression of EOR during surgery and postoperative imaging measurements.[15,16] Initial studies on the accuracy of neurosurgeons' estimates of resectability are limited to glioblastoma[16–18]. This has not been addressed for nonenhancing glioma.

Several means to support the interpretation of resectability have been proposed, such as the anatomical landmarks of functional brain regions,[8,19–21] an atlas of brain functions,[22,23] specific radiological features,[11,24,25] or automated tumor segmentation image analysis.[26] Another source of potentially useful information is resection probability maps (RPMs) based

on a large number of prior resections of nonenhancing glioma normalized to standard brain space. For low-grade glioma, RPMs have been previously used to estimate the expected residual tumor volume,[27] and to evaluate the potential for brain plasticity,[22] and to compare resection results between surgical teams.[28]

Here, we determined the accuracy and the precision of preoperative estimates of resectability by neurosurgeons with more and less experience in glioma surgery, and compared these with estimates of an RPM.

METHODS

Patients

An observational cohort of consecutive patients was identified from two tertiary referral centers for neuro-oncological surgery (Montpellier and Amsterdam). From each center patients were included, who (1) were over 17 years of age, (2) were diagnosed with a supratentorial glioma of intermediate grade

(WHO II or III with focal anaplasia[1]), (3) had a resection between 2006 and 2012, (4) had T2/FLAIR hyperintense abnormalities as target for resection, (5) had no prior radiotherapy to avoid misinterpretation of T2/FLAIR hyperintensity, (6) had a 3- to 6-month postoperative MRI available.

Patients received standard care and treatment decisions were made in tumor board meetings. Neurosurgical teams performed resections up to functional boundaries by applying intraoperative stimulation mapping under local anesthesia. A subset of the data (108 patients) was analyzed previously to demonstrate similar extent of resection between the two surgical teams.[28]

The institutional review board approved the study protocol and informed consent was obtained from patients. Data for analysis was anonymized.

MR-Scanning

Patients underwent a pre- and postoperative MR-scan using a standardized protocol. For analysis, MRIs scheduled at four months postoperatively were used to avoid surgical artifacts. Imaging was performed on various 1.5 and 3.0 T scanners including Siemens Avanto 1.5 T (Siemens Healthcare GmbH, Erlangen Germany), GE Signa HDXT 1.5 and 3.0 T (General Electeric HealthCare, Chicago, Illinois) and Toshiba Titan 3.0 T (Canon Meical Systems Inc,

Tustin, California). Sagittal 3D turbo T2 fluid-attenuated inversion-recovery (FLAIR) images (repetition time/echo time/inversion time 6500/335/2000-2200 for 1.5 T and 6000-6500/126-401,5/2000 for 3.0 T) with 1.3-2.0 mm section thickness, and 3-mm reconstructions in axial plane (oriented along the hypophysis–fastigium line) were derived. The protocol included a 3D heavily T1-weighted sequence, obtained after administration of intravenous gadolinium.

Observed Resection Results

The observed EOR, 100% x (1 - residual tumor volume/preoperative tumor volume), and observed residual volume in mL were determined from the segmentations of pre- and postoperative MRIs and used for further analysis.

Resectability Estimates by Neurosurgeons

Six neurosurgeons estimated the expected EOR as percentage and expected residual volume in mL by examining the anonymized preoperative MRIs consisting of T2/FLAIR- and T1-weighted images of every patient, while blinded for postoperative resection results and outcome. Age, dexterity and information on preoperative neurological examination were provided to simulate the setup during tumor board meetings.

The level of experience with glioma surgery was either senior or junior for the neurosurgeons, one of each from teams from No rater had been involved in the decision-making or surgery of these patients. The senior raters had 30, 24, and 15 years of experience in glioma surgery. The junior raters were chief-resident or in their first year following residency completion.

Resectability Estimates Based on the Resection Probability Map

The processing of the RPMs has been described previously[28] with some minor updates in the methods as detailed in the Supplemental Digital Content.

Resectability estimates of the preoperative images could be derived from this RPM. For each patient, an estimate of expected residual volume in mL was calculated by summing the resection probabilities of the voxels containing tumor for that patient. The expected EOR as percentage was then calculated accordingly, 100% x (1 - residual volume / preoperative tumor volume).

A leave-one-out approach was used to estimate the resectability for each patient: the observation was excluded from the RPM used for the estimate of that patient.

Statistical Analysis

We approached the analysis as diagnostic method comparison problem with paired observations using the observed resection results, estimated resectability by senior and junior neurosurgeons and the RPM, with three exchangeable replicates for junior and senior neurosurgeons.[29]

A minimum sample size of 203 patients was calculated based on the assumptions of a target intra-class coefficient of 0.75, three replicates per neurosurgical expertise level, a confidence level of 0.95, an expected interval width of 0.1.[30]

As exchangeable replicates had to be accommodated, we chose to adopt a hierarchical Bayesian framework to estimate the bias, the limits of agreement and the intraclass correlation from the observed and estimated EORs and residual tumor volumes.[31] Vague priors with zero mean and low precision (0.0001) were chosen to primarily reflect inference from the presented data without substantive prior knowledge. The mean values of posterior distributions were used as estimates with 95% credibility intervals (95% CrI) based on three parallel chains of 5,000 Markov Chain Monte Carlo simulation samples after the same number of burn-ins in OpenBUGS.[32] The model is provided in the Supplemental Digital Content. No evidence against convergence was identified using sampling traces and

distributions and Gelman-Rubin diagnostics.

First, we made boxplots of the observed and estimated extents of resection and residual volumes for each method. Second, we created scatterplots of the estimated and observed extents of resection and residual volumes per method. Third, we created Bland–Altman plots for the agreement between the estimations and the observed resections, as reference.[29,33] We calculated the bias with 95% CrI to determine the accuracy of the estimates per method, the limits of agreement with 95% CrI to evaluate the precision of the estimates per method, and the intra-class correlation coefficient to assess the consistency of measurements with replicates, i.e. junior and senior neurosurgeons.[31] We considered estimates biased for a method when zero was excluded from the 95% CrI. The intraclass correlation was interpreted as consistency between estimates.[34] Finally, as it is more important to classify patients as good or poor candidates for resective surgery, we applied a Receiver Operating Characteristic (ROC) analysis to calculate the diagnostic accuracy of the three methods to predict the observed result. The areas under the curve of the ROC curves were compared using the bootstrap method.[35]

RESULTS

Patient and Tumor Characteristics

Our observational patient cohort consisted of 234 adults with nonenhancing diffuse glioma. The median age was 39.4 years (interquantile range [IQR] of 31.5 to 47.1), 103 (44%) patients were female. Of all patients 150 (64%) had resective surgery in Montpellier, 84 (36%) in Amsterdam. None of these patients had new permanent neurological deficits.
The gliomas had a median volume of 50 mL (IQR: 26 to 81) and were located in the left hemisphere in 130 (56%). The median postoperative residual volume was 8.8 mL (IQR: 1.8 to 21), and the median EOR was 82% (IQR: 62 to 96%). Less than 20 mL residual tumor was observed in 173 (%74) patients, less than 10 mL in 128 (55%), and less than 5 mL in 92 (39%). More than 40% EOR was observed in 213 (91%) patients, more than 70% in 160 (68%), and more than 90% in 87 (37%). The RPM based on all patients is plotted in Figure 1, and Video 1 and 2.
Illustrative cases for resection estimates are shown in Figure 2.

Estimates of Extent of Resection

Estimates were within 10% of the observed EOR in 101 (43%) observations for the RPM, 101 (43%), 68 (29%), and 113 (48%) observations for the three junior neurosurgeons, and 111 (47%), 114 (49%), and 115 (49%) observations for the three senior neurosurgeons. Two junior neurosurgeons had one missing estimate for EOR, one senior neurosurgeon had two missing estimates, and another senior neurosurgeon had one missing estimate.
The observed and estimated extents of resection from the RPM, and junior and senior neurosurgeons are displayed in Figure 3. Both junior and senior neurosurgeons estimated the EOR somewhat but systematically higher than the observed values, except for one junior

neurosurgeon (j2), who systematically underestimated with a wide distribution (Figure 3A). The RPM estimates had a distribution similar to the observed values.

The tendency to overestimate by neurosurgeons was confirmed by a plot of the EOR values of the RPM, junior and senior neurosurgeons against the observed values with a high density of values at the right side of the plot (Figure 3B) corresponding to estimates that are larger than the observed EOR.

Estimates of Residual Tumor Volume

Estimates were within 5 mL of the observed residual tumor volume in 119 (51%) observations for the RPM, 104 (44%), 85 (36%), and 107 (46%) observations for the junior neurosurgeons, and 115 (49%), 110 (47%), and 111 (47%) observations for the senior neurosurgeons.

The observed and estimated residual tumor volumes from the RPM, and junior and senior neurosurgeons are displayed in Figure 4. Both junior and senior neurosurgeons estimated residual tumor volumes somewhat lower than the observed values (Figure 4A).

The tendency to underestimate the residual tumor volumes by neurosurgeons is seen as high density of values at the left side of the plots in Figure 4B, indicating estimates to be smaller than the observed residual tumor volumes. The RPM estimates are distributed more evenly above and below the diagonal.

Bland-Altman Analysis of Extent of Resection

The Bland-Altman plots in Figure 3C demonstrate that the accuracy of estimates depends on the complexity of the gliomas and the experience of neurosurgeons. For instance, the least complex gliomas with a large EOR of over 80% were estimated more accurately than the gliomas with an EOR of less than 80% (Figure 3C). The systematic overestimation of EOR

was quantitated by the bias, which was 4.2% (95% CrI: 0.9 to 7.5%) by junior neurosurgeons and 11.2% (95% CrI: 8.3 to 14%) by senior neurosurgeons. The RPM provided unbiased estimates, i.e. a bias of -2.6% (95% CrI: -5.8 to 0.5%). The 95% limits of agreement for EOR were wide for all methods, but wider for the junior neurosurgeons than for the seniors.: -50 to 46% for EOR by the RPM, -58% to 68% by junior neurosurgeons, and -37% to 59% by senior neurosurgeons. The intraclass correlation of the estimated EORs was lower among junior neurosurgeons (0.18, 95% CrI: 0.13 to 0.24) than among senior neurosurgeons (0.36, 95% CrI: 0.29 to 0.44), indicating poor consistency.

Bland-Altman Analysis of Residual Tumor Volume

For residual tumor volume, the Bland- Altman plots in Figure 4C demonstrate that the accuracy of estimates depends on the complexity of the gliomas and the experience of neurosurgeons. Less complex gliomas with regard to small residual volumes were estimated more accurately than large residual volumes which were typically underestimated by neurosurgeons (Figure 4C). The systematic underestimation of residual tumor volume was also demonstrated by the bias of -4.3 mL (95% CrI: -6.8 to -1.8) for junior neurosurgeons and -9.0 mL (95% CrI: -11 to -7.0) for senior neurosurgeons. On the contrary, estimates of residual tumor volume by RPMs were unbiased: 0.2 mL (95% CrI: -1.9 to 2.2). The 95% limits of agreement for residual tumor volume were wide for all methods, but wider for the junior neurosurgeons than for the seniors: -32 to 32 mL for residual volume by the RPM, -51 to 44 mL by junior neurosurgeons, and -43 to 25 mL by senior neurosurgeons. The intraclass correlation of the estimated residual tumor volumes was lower among junior neurosurgeons (0.12, 95% CrI: 0.07 to 0.17) than among senior neurosurgeons (0.28, 95% CrI: 0.22 to 0.34), indicating poor consistency.

Receiver Operating Characteristics Analysis of Extent of Resection

From the ROC plots in Figure 3D, the RPM was more accurate in identifying the patients in whom at least 40% EOR was obtained with an area under the curve (auc) of 0.76 compared with 0.61 for junior neurosurgeons (P=0.01) and compared with 0.60 for senior neurosurgeons (P=0.008). No significant difference in diagnostic accuracy was observed in identifying patients with at least 70% or 90% EOR between the RPM (auc: 0.69 and 0.75, respectively), the junior neurosurgeons (auc: 0.63 and 0.69, respectively), and the senior neurosurgeons (auc: 0.70 and 0.76, respectively).

Receiver Operating Characteristics Analysis of Residual Tumor Volume

In the ROC plots in Figure 4D, the RPM was as accurate as the senior neurosurgeons to identify patients in whom to expect a residual tumor volume of less than 5, 10 or 20 mL. The RPM was significantly more accurate than the junior neurosurgeons to identify the patients in whom to expect a residual tumor volume of less than 5 mL (auc: 0.84 and 0.75, P=0.009), 10 mL (auc: 0.84 and 0.73, P=0.001) and 20 mL (auc: 0.84 and 0.71, P=0.0004). The RPM estimates were as accurate as these from the senior neurosurgeons (auc: 0.83, 0.81, 0.81, not significant).

DISCUSSION

The main findings of our study are that neurosurgeons estimate resectability before surgery rather accurate with small but systematic overestimation, whereas a RPM provides accurate unbiased estimates. The RPM estimates may be particularly advantageous in identifying the patients in whom a resection of more than 40% can be obtained.

Where quantitative volumetric MRI analysis has superseded the intraoperative impression of resection completeness in glioma surgery, a quantitative strategy should improve preoperative interpretation of resectability. The preoperative subjective interpretation by neurosurgeons has been demonstrated to be inaccurate in glioblastoma[16-18], but for nonenhancing glioma this has not been determined. Recently, Sonabend et al[36] demonstrated varying resectability estimates between 13 neurosurgeons on 20 glioblastoma patients, but the resectability index calculated from pooled responses strongly correlated with the observed residual tumor ($R=0.817$, $P<0.001$). Resection indications should improve from unbiased resectability estimates, if an evidence-based minimum threshold for a useful resection would be available for nonenhancing glioma. A resection threshold has however not been established. Reports on surgical outcome vary widely in classification of resection completeness with cutoffs between 40% to 100% resection and between 0 to 30 mL of residual tumor.[4,11–13,37–39] The concept of a threshold, below which a resection would be futile and above which it would be worthwhile, is possibly an oversimplication of what is more likely to be a gradual increase in survival benefit from less residual disease.[40] Furthermore, a partial resection without survival benefit may reduce symptoms from mass effect or control seizures. Nevertheless, the RPM not only provides an overall resectability estimate for the individual patient, but also allows for a quantitative overlay over the tumor to visualize which regions are more or less likely to be resected, as demonstrated in Figure 2. This could serve as an adjunct to preoperative

surgical planning.

An alternative explanation for the discrepancy between neurosurgeons' estimates and observed resection results can be that the estimating neurosurgeons had a strict oncological perspective on resectability, blinded for patient input on what was considered acceptable functional outcome. With this input the treating neurosurgeons could have been less aggressive in pushing the resection to the limits of resectability in some patients. This is supported by the fact that no patients were excluded from this cohort due to new permanent neurological deficits. For instance, some patients with a glioma in the supplementary motor area accept a permanent subtle sensomotor decline for maximizing tumor removal, because it does not interfere with their activities. Other patients prefer to have risks minimized for continuation of their activities at high level, such as sports, dance, or playing an instrument. In other words, surgical decisions may be made differently when the oncological goal of tumor removal is balanced with the functional goal of preservation of functional integrity in shared decision-making based on the patient's values, sometimes referred to as the 'oncofunctional balance'.[41,42]

The limits of agreement can be interpreted as a measure to compare the total difference between an estimation method and a reference. This difference includes both systematic bias (accuracy) and random error (precision). In our results the 95% limits of agreement between either neurosurgeons or the RPM and the observed resection results are wider than what can be considered clinically useful. This seems to be more prominent in more complex gliomas with an observed EOR less than 80%, whereas in less complex gliomas estimations seem to be more useful. Furthermore, the intraclass coefficients for resection predictions between junior and senior neurosurgeons are low. Apparently, other factors than age, handedness, symptoms, and tumor location were involved in the surgical decision-making in these cases and a potential source of error. We speculate that anticipated shorter duration of postoperative

recovery may result in resections that are not in contact with the functional brain structures, guided by patient preferences and/or the urge for undelayed adjuvant treatment. Therefore, resectability estimates using a RPM cannot replace clinical judgment by neurosurgeons, but should rather be considered as a supportive instrument.

Instead of static rules for inoperable brain regions,[20] sometimes referred to as eloquence, the RPM may provide a dynamic quantitative map that captures current surgical decision-making using modern techniques. [23,43,44] As such, this can be an educational instrument for junior neurosurgeons, and a feedback instrument for senior neurosurgeons. For instance, junior neurosurgeons had widely varying precision in their resection predictions, despite a similar number of years in training in comparable university hospitals with brain tumor programs. The junior neurosurgeons may have had different exposure to advanced techniques in glioma surgery. Their predictions may improve from the information in an RPM.

Strengths

The strengths of this study include involvement of a relatively large number of neurosurgeons as raters with two distinct levels of experience from different countries covering the diversity of clinical practice, blinded for surgical outcome. Furthermore, we report on a large number of unselected patients with nonenhancing glioma, who had resective surgery by two different teams from the same school with similar outcomes.[28] Lastly, we adhered to standards for reporting on method comparison studies according to guidelines.[45]

Limitations

The limitations of this study are that manual tumor segmentation and brain registration can be subject to variation and are therefore potential sources of error in our data, although visual inspection verified satisfactory processing in all patients. Furthermore, we did not determine

the intrarater agreement in our analysis, as we rather invested the time resources of the raters in a larger number of patients than to have repeated estimates on the same patients with uncertain blinding to previous estimates.

Future efforts could focus on the development of software applications to facilitate image processing and analysis available to the neuro-oncological community. In addition, the RPM is a dynamic representation of current surgical results, that can be updated or improved by contributing new patients possibly having had surgery with other techniques, rather than being a static universal characteristic of brain regions. Finally, the certainty of resectability estimates varies throughout the brain, depending on the preferential locations of nonenhancing gliomas.

CONCLUSION

Junior and senior neurosurgeon estimates of resectability before surgery of a nonenhancing glioma were rather accurate - with a small but systematic bias - and imprecise - with wide limits of agreement. An RPM provides unbiased resectability estimates, which can be useful for surgical decision-making, planning, and education.

REFERENCES

1. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. Acta neuropathol. 2007;114(2):97-109.

2. Soffietti R, Baumert B, Bello L, et al. Guidelines on management of low-grade gliomas: report of an EFNS-EANO Task Force. Eur J Neurol. 2010;17(9):1124-1133.

3. Weller M, Bent M van den, Tonn JC, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol. 2017;18(6):e315-e329.

4. 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.

5. Sanai N, Polley M-Y, McDermott MW, Parsa AT, Berger MS. An extent of resection threshold for newly diagnosed glioblastomas. J Neurosurg. 2011;115(1):3-8.

6. Jakola AS, Myrmel KS, Kloster R, et al. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. JAMA. 2012;308(18):1881-1888.

7. Aghi MK, Nahed BV, Sloan AE, Ryken TC, Kalkanis SN, Olson JJ. The role of surgery in the management of patients with diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. J Neurooncol. 2015;125(3):503-530.

8. 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.

9. 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.

10. Oppenlander ME, Wolf AB, Snyder LA, et al. An extent of resection threshold for

recurrent glioblastoma and its risk for neurological morbidity. J Neurosurg. 2014;120(4):846-853.

11. Ius T, Isola M, Budai R, et al. Low-grade glioma surgery in eloquent areas: volumetric analysis of extent of resection and its impact on overall survival. A single-institution experience in 190 patients. J Neurosurg. 2012;117(December):1-14.

12. Capelle L, Fontaine D, Mandonnet E, et al. Spontaneous and therapeutic prognostic factors in adult hemispheric World Health Organization Grade II gliomas: a series of 1097 cases: clinical article. J Neurosurg. 2013;118(6):1157-1168.

13. Roelz R, Strohmaier D, Jabbarli R, et al. Residual Tumor Volume as Best Outcome Predictor in Low Grade Glioma - A Nine-Years Near-Randomized Survey of Surgery vs. Biopsy. Sci Rep. 2016;6(1):32286.

14. Xu DS, Awad A-W, Mehalechko C, et al. An extent of resection threshold for seizure freedom in patients with low-grade gliomas. J Neurosurg. May 2017:1-7.

15. Albert FK, Forsting M, Sartor K, Adams HP, Kunze S. Early postoperative magnetic resonance imaging after resection of malignant glioma: objective evaluation of residual tumor and its influence on regrowth and prognosis. Neurosurgery. 1994;34(1):45-60; discussion 60-61. 16. Orringer D, Lau D, Khatri S, et al. Extent of resection in patients with glioblastoma: limiting factors, perception of resectability, and effect on survival. J Neurosurg. 2012;117(5):851-859. 17. Coburger J, Renovanz M, Ganslandt O, Ringel F, Wirtz CR, Riehm JS von. 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.

18. Capellades J, Teixidor P, Villalba G, et al. Results of a multicenter survey showing interindividual variability among neurosurgeons when deciding on the radicality of surgical resection in glioblastoma highlight the need for more objective guidelines. Clin Transl Oncol.

2017;19(6):727-734.

19. 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. 2001;52(2):161-171.

20. 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):1044-1055.

21. Sanai N, Polley M-Y, Berger MS. Insular glioma resection: assessment of patient morbidity, survival, and tumor progression. J Neurosurg. 2010;112(1):1-9.

22. Ius T, Angelini E, Thiebaut de Schotten M, Mandonnet E, Duffau H. Evidence for potentials and limitations of brain plasticity using an atlas of functional resectability of WHO grade II gliomas: Towards a “minimal common brain”. Neuroimage. 2011;56(3):992-1000.

23. 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.

24. 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.

25. Baldock AL, Ahn S, Rockne R, et al. Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas. PloS one. 2014;9(10):e99057.

26. Meier R, Porz N, Knecht U, et al. Automatic estimation of extent of resection and residual tumor volume of patients with glioblastoma. J Neurosurg. 2017;121(5):1-9.

27. 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.

28. 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.

29. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135-160.

30. Shieh G. Sample size requirements for the design of reliability studies: precision consideration. Behav Res Methods. 2014;46(3):808-822.

31. Schluter PJ. A multivariate hierarchical Bayesian approach to measuring agreement in repeated measurement method comparison studies. BMC Med Res Methodol. 2009;9(1):6.

32. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. Stat Med. 2009;28(25):3049-3067.

33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-310.

34. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284-290.

35. 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(1):77.

36. 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.

37. Youland RS, Schomas DA, Brown PD, et al. Changes in presentation, treatment, and outcomes of adult low-grade gliomas over the past fifty years. Neuro Oncol. 2013;15(8):1102- 1110.

38. McGirt MJ, Chaichana KL, Attenello FJ, et al. Extent of surgical resection is independently associated with survival in patients with hemispheric infiltrating low-grade gliomas. Neurosurgery. 2008;63(4):700-7; author reply 707-708.

39. Pignatti F, Bent M van den, Curran D, et al. Prognostic factors for survival in adult patients with cerebral low-grade glioma. J Clin Oncol : official journal of the American Society of Clinical Oncology. 2002;20(8):2076-2084.

40. Solheim O, Gulati S, Jakola AS. Glioblastoma resection: in search of a threshold between worthwhile and futile. Neuro Oncol. 2014;16(4):610-611.

41. Duffau H, Mandonnet E. The “onco-functional balance” in surgery for diffuse low-grade glioma: integrating the extent of resection with quality of life. Acta Neurochir. 2013;155(6):951- 957.

42. Kane HL, Halpern MT, Squiers LB, Treiman KA, McCormack LA. Implementing and Evaluating Shared Decision Making in Oncology Practice. CA Cancer J Clin. 2014;64(6):377- 388.

43. Sanai N, Mirzadeh Z, Berger MS. Functional Outcome after Language Mapping for Glioma Resection. N Engl J Med. 2008;358(1):18-27.

44. Herbet G, Maheu M, Costi E, Lafargue G, Duffau H. Mapping neuroplastic potential in brain-damaged patients. Brain. 2016;139(3):829-844.

45. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106.

FIGURE LEGENDS

Figure 1. The RPM based on nonenhancing gliomas of 130 patients with a tumor in the left hemisphere in the upper row and 104 patients in the right hemisphere in the lower row. Separate RPMs were used for gliomas in the left and right hemisphere to facilitate resectability estimates in contralateral brain regions. The numbers correspond with the Montreal Neurological Institute brain template z-axis coordinates. In the color legend green corresponds to locations with a resectability index of 1.0 where in all patients with a tumor at that location the tumor was removed, and red corresponds with locations with a resectability index of 0.0 where in none of the patients the tumor was removed. The RPM is available as movie for left-sided and right-sided tumors (see Videos 1 and 2, respectively).

Figure 2. Two examples of pre- and postoperative FLAIR-weighted imaging with resectability based on the RPM. A, A 40-year old right-handed male had an incidental finding of a 23 mL nonenhancing glioma. The estimated EOR was 100%, 100%, and 95% by junior neurosurgeons, 100%, 100%, and 99% by the senior neurosurgeons, and 95% by the RPM. The estimated residual volume was 0, 0, and 1 mL by junior neurosurgeons, 0, 0, and 0.2 mL by the senior neurosurgeons, and 1 mL by the RPM. The observed EOR and residual tumor volume were 100% and 0 mL. B, A 41-year old right-handed male presented with a seizure and a 36 mL nonenhancing glioma. The estimated EOR was 100%, 100%, and 95% by junior neurosurgeons, 100%, 98%, and 96% by the senior neurosurgeons, and 63% by the RPM. The estimated residual volume was 0 mL, 0 mL, and 1.2 mL by the junior neurosurgeons, 0 mL, 0.5 mL, and 2 mL by the senior neurosurgeons, and 16 mL by the RPM. The observed EOR and residual tumor volume were 61% and 14 mL.

Figure 3. Agreement between observed resections (yellow) and estimates of EOR from the RPM (blue), junior neurosurgeons (green), and senior neurosurgeons (red). A, Boxplots of the

234 observations and estimates to display the distributions with median, 25% and 75% quartiles as hinges and 1.5 times the interquartile distance as whiskers, with outliers. B, Scatterplot of estimated versus observed extents of resection. Each dot represents an estimated and observed EOR for one patient. Individual junior and senior neurosurgeons are plotted as circles, squares and triangles. Datapoints are partially transparent to demonstrate high densities of data. C, Bland-Altman plots of the mean of estimated and observed extents of resection versus the difference of expected minus observed extents of resection. Each dot represents one patient. Estimates from individual junior and senior neurosurgeons are plotted as circles, squares and triangles. The bias is plotted as solid line with 95% CI as dotted lines. The limits of agreement are plotted as dashed lines. D, The receiver operating characteristic curves to identify patients with more than 40%, 70%, and 90% EOR, respectively, for estimates by the RPM in blue, junior neurosurgeons in green, and senior neurosurgeons in red.

Figure 4. Agreement between observed resections (yellow) and estimates of residual tumor volume from the RPM (blue), junior neurosurgeons (green), and senior neurosurgeons (red). A, Boxplots of the 234 observations and estimates to display the distributions with median, 25% and 75% quartiles as hinges and 1.5 times the interquartile distance as whiskers, with outliers. B, Scatterplot of estimated versus observed residual tumor volumes. Each dot represents an estimated and observed EOR for one patient. Individual junior and senior neurosurgeons are plotted as circles, squares and triangles. Datapoints are partially transparent to demonstrate high densities of data. C, Bland-Altman plots of the mean of estimated and observed residual tumor volumes versus the difference of expected minus observed residual tumor volumes. Each dot represents one patient. Estimates from individual junior and senior neurosurgeons are plotted as circles, squares and triangles. The bias is plotted as solid line with 95% CI as dotted lines. The limits of agreement are plotted as dashed lines. D, The

receiver operating characteristic curves to identify patients with less than 5mL, 10mL, and 20mL residual tumor volume, respectively, for estimates by the RPM in blue, junior neurosurgeons in green, and senior neurosurgeons in red.

VIDEO LEGENDS

Video 1. The RPM based on nonenhancing gliomas of 130 patients with a tumor in the left hemisphere. The numbers below correspond with the Montreal Neurological Institute brain template z-axis coordinates. In the color legend green corresponds to locations with a resectability index of 1.0 where in all patients with a tumor at that location the tumor was removed, and red corresponds with locations with a resectability index of 0.0 where in none of the patients the tumor was removed.

Video 2. The RPM based on nonenhancing gliomas of 104 patients with a tumor in the right hemisphere. Montreal Neurological Institute brain template z-axis coordinates and color legend are identical as defined in Video 1.

SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content. Details of the RPM processing and the Bayesian model code.

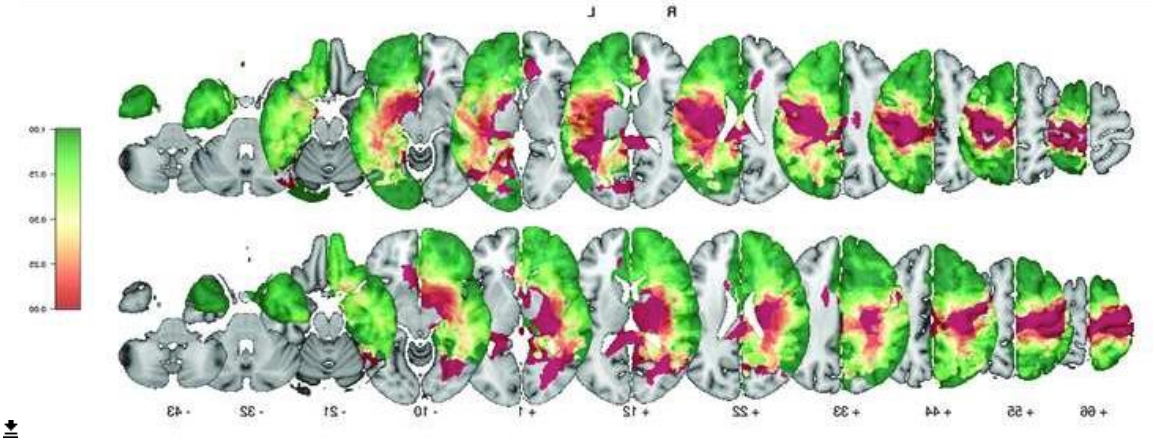

Figure 1 Click here to access/download;Figure;FIG_1.tif

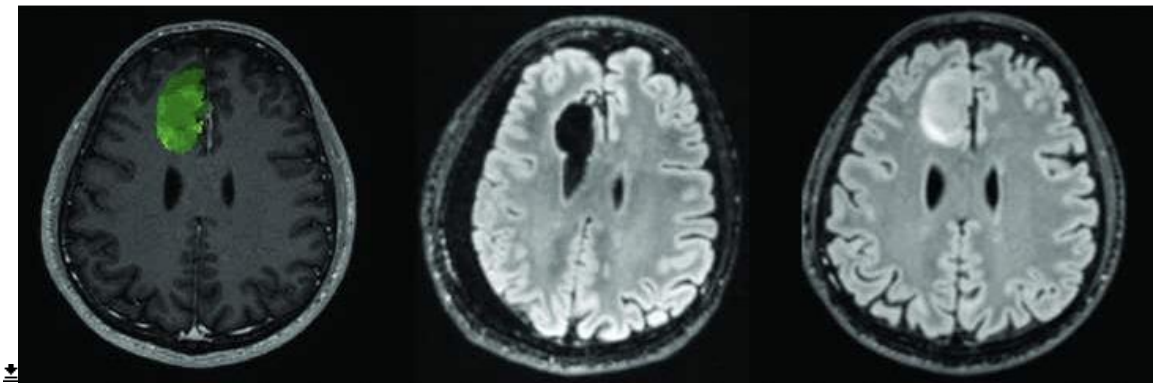

Figure 2A Click here to access/download;Figure;FIG_2A.tif

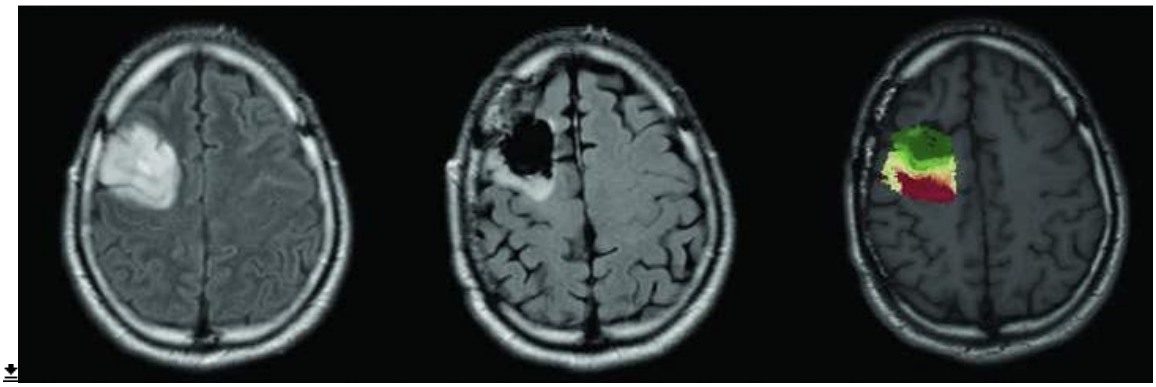

Figure 2B Click here to access/download;Figure;FIG_2B.tif

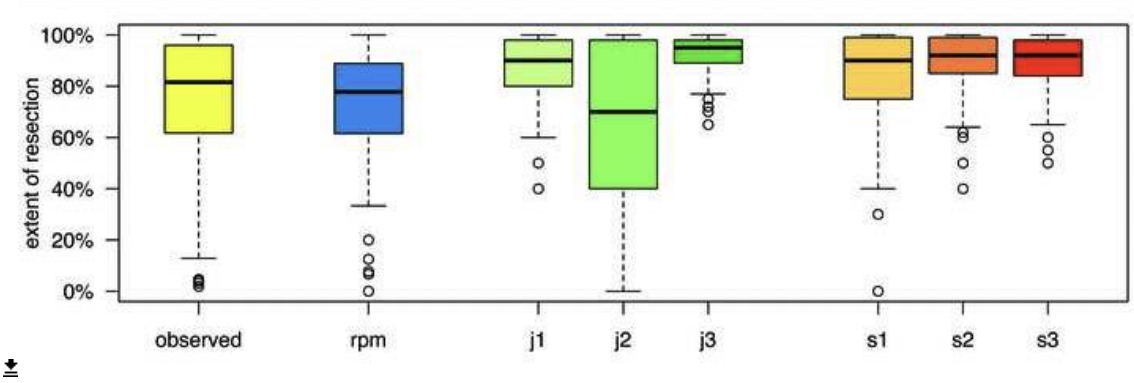

Figure 3A Click here to access/download;Figure;FIG_3A.tif

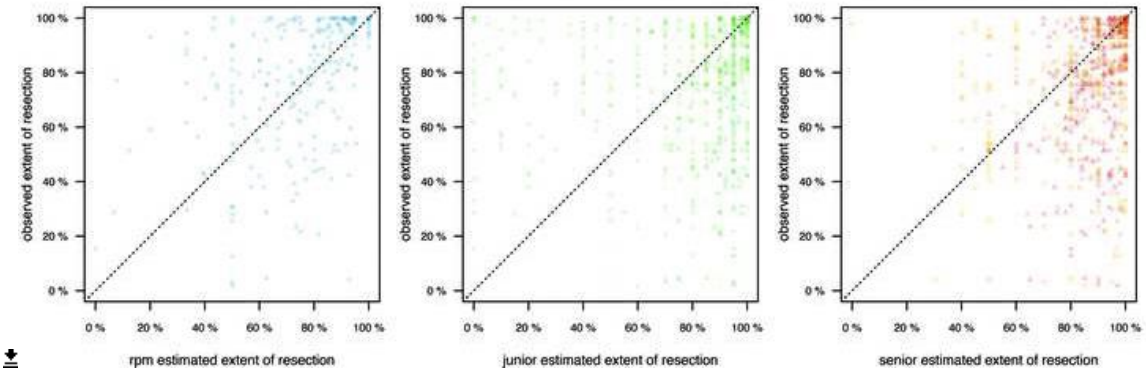

Figure 3B Click here to access/download;Figure;FIG_3B.tif

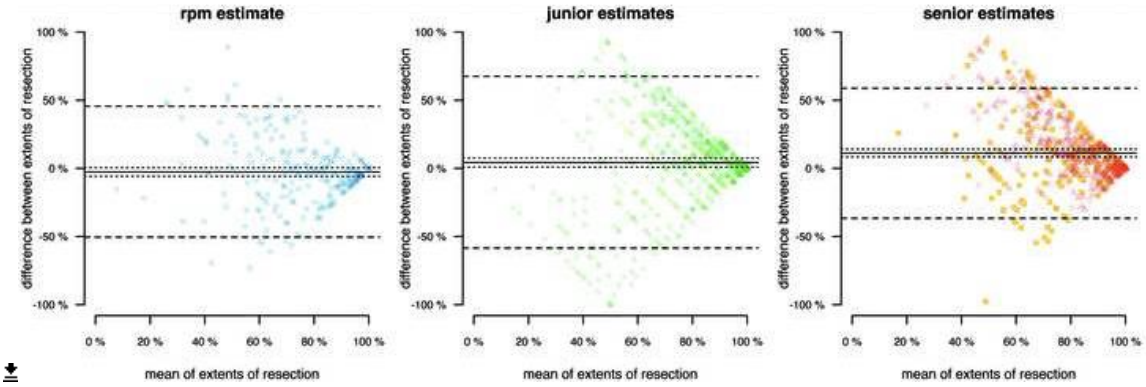

Figure 3C Click here to access/download;Figure;FIG_3C.tif

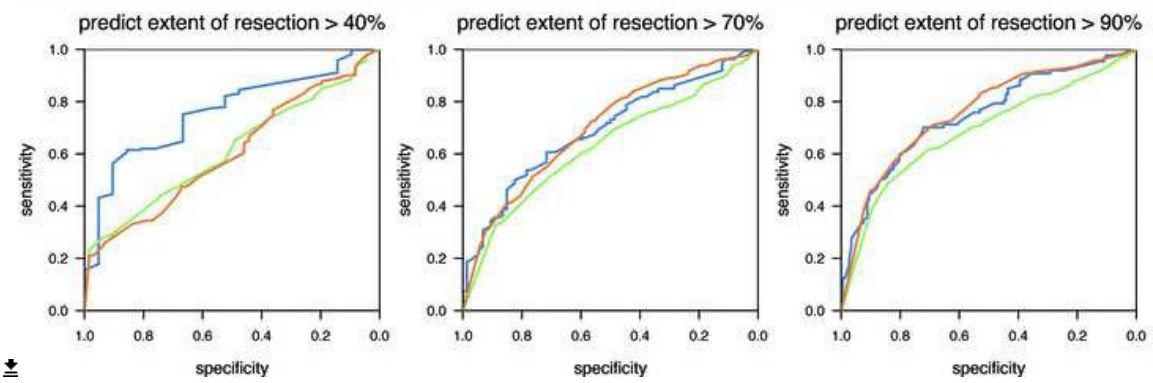

Figure 3D Click here to access/download;Figure;FIG_3D.tif

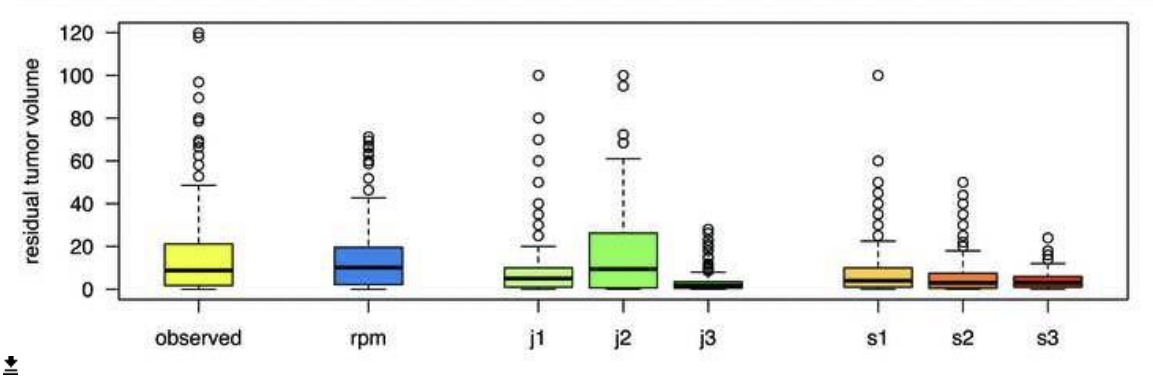

Figure 4A Click here to access/download;Figure;FIG_4A.tif

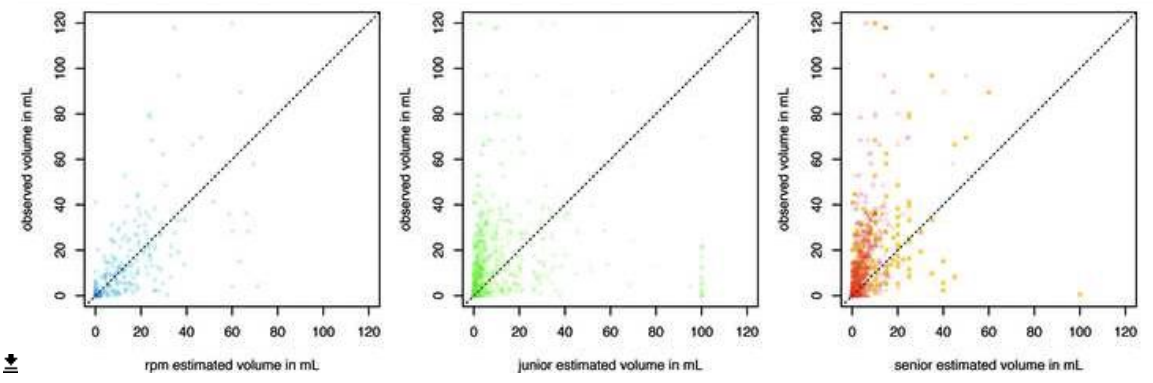

Figure 4B Click here to access/download;Figure;FIG_4B.tif

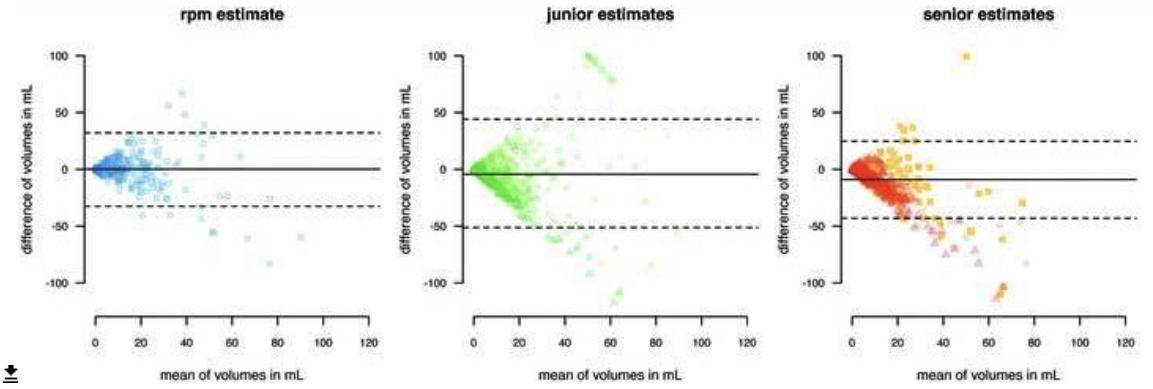

Figure 4C Click here to access/download;Figure;FIG_4C.tif

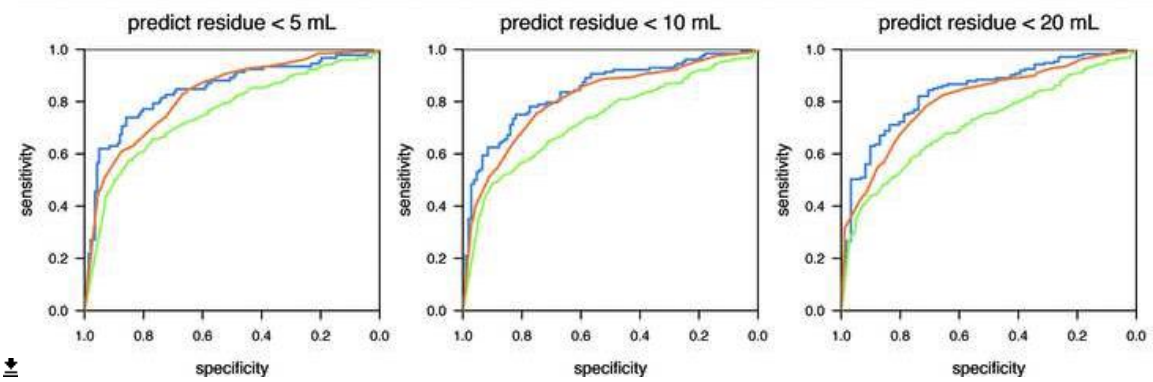

Figure 4D Click here to access/download;Figure;FIG_4D.tif

Supplemental Digital Content.

Processing of the RPM

The resection regions were segmented in 3D on the postoperative FLAIR images using the smart brush tool of iPlan v3.0 software (BrainLAB AG, Feldkirchen, Germany). The segmented volumes were verified and adjusted in reconstruction planes by two observers (E.H. and P.W.H.). To ensure correct interpretation of T2/FLAIR-hyperintensity on the MRIs four months after surgery, the MRIs before surgery and days after surgery (i.e. within 72 hours) with diffusion-weighted images were taken into account to avoid inadvertent inclusion of postoperative contusion, gliosis or ischemia as tumor. None of the patients had cerebrovascular pathology or white matter disease. The T2/FLAIR-weighted images were linearly registered with the T1-weighted images for further analysis using the image fusion tool of iPlan v3.0 software (BrainLAB AG, Feldkirchen, Germany). The segmented volumes were exported as binary masks registered to the T1-weighted images. To correlate these objects in standard brain space, i.e. the 152 T1 normal brain template of 1 mm available from the Montreal Neurological Institute (MNI) the T1-weighted images were non-linearly registered in sequential steps, including rigid, affine, B-spline regularization and symmetric diffeomorphic registration with cross-correlation as similarity metric. The segmented objects were then aligned as binary images to standard brain space using these patient-specific non-linear transformations, resulting in preoperative tumor and postoperative residual volumes of all patients that were aligned in 1x1x1 mm resolution for further analysis. Subsequently, at each of the approximately 1.5 million voxels that cover the standard brain space, the probability of resection was calculated for each voxel by dividing the number of patients without residual tumor by the summary of patients having a glioma at that voxel, as

previously described. This resulted in a high-resolution RPM for likelihood of resection within the brain for the entire patient cohort.

Hierarchical Bayesian model

The OpenBUGS model for diagnostic method comparison with paired observations using the observed resection results, estimated resectability by senior and junior neurosurgeons and the RPM, with three exchangeable replicates for junior and senior neurosurgeons was adopted from Schluter et al (BMC Medical Research Methodology 2009, 9:6).

```
 model{

  for (i in 1:Subjects){

for (j in 1:Replications){
x[i,j,1:Methods] ~ dmnorm(mu[i,1:Methods],P[,])

}

mu[i,1:Methods] ~ dmnorm(theta[1:Methods],T[,]) }

for (k in 1:Methods){
theta[k] ~ dnorm(0,0.0001)
# Intra-classcorrelation within method with replicates
ICC[k] <- V[k,k]/(V[k,k]+W[k,k])
dV[k] <- V[k,k] # diagonal V # sigma^2: covariance between subjects dW[k] <- W[k,k] #
diagonal W # sigma^2: covariance between methods

}
P[1:Methods,1:Methods] ~ dwish(U[1:Methods,1:Methods],Methods)
T[1:Methods,1:Methods] ~ dwish(R[1:Methods,1:Methods],Methods)
W[1:Methods,1:Methods] <- inverse(P[1:Methods,1:Methods]) V[1:Methods,1:Methods] <-
inverse(T[1:Methods,1:Methods]) p.mu[1:Methods] ~ dmnorm(theta[1:Methods],T[,]) #
```

```
Predicted means p.x[1:Methods] ~ dmnorm(p.mu[1:Methods],P[,]) # Predicted measurements
for (y in 1:1){

for (z in 1:Methods){
bias[y,z] <- theta[y] - theta[z] # Bias estimation diff[y,z] <- p.x[y] - p.x[z] # Limits of
agreement

} }

  # test difference between bias rpm and junior

  dbias[1] <- bias[1,2] - bias[1,3]

  # test difference between bias rpm and senior

  dbias[2] <- bias[1,2] - bias[1,4]

}
```