Adam M. Sonabend, MD*
Brad E. Zacharia, MD, MS‡
Michael B. Cloney, BA*
Aarón Sonabend, PhD§
Christopher Showers, BA*
Victoria Ebiana, MD*
Matthew Nazarian, BS*
Kristin R. Swanson, PhD¶
Anne Baldock, BA‖ †
Henry Brem, MD#
Jeffrey N. Bruce, MD*
William Butler, MD**
Daniel P. Cahill, MD, PhD**
Bob Carter, MD, PhD‡‡
Daniel A. Orringer, MD§§
David W. Roberts, MD¶¶
Oren Sagher, MD§§
Nader Sanai, MD‖‖
Theodore H. Schwartz, MD##
Daniel L. Silbergeld, MD***
Michael B. Sisti, MD*
Reid C. Thompson, MD‡‡‡
Allen E. Waziri, MD§§§
Guy McKhann, II, MD*

*Department of Neurological Surgery, College of Physicians and Surgeons, Columbia University Medical Center, New York, New York; ‡Department of Neurosurgery, Penn State Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, Pennsylvania; §Instituto Tecnologico de Mexico, Department of Applied Mathematics, Mexico City, Mexico; ¶Department of Neurological Surgery, Mayo Clinic, Scottsdale, Arizona; ‖University California at San Diego School of Medicine, San Diego, California; #Department of Neurological Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland; **Department of Neurological Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ‡‡Division of Neurosurgery, Department of Surgery, University California at San Diego School of Medicine, San Diego, California; §§Department of Neurological Surgery, University of Michigan School of Medicine, Ann Arbor, Michigan; ¶¶Division of Neurosurgery, Dartmouth University, Lebanon, New Hampshire; ‖‖Division of Neurosurgical Oncology, Barrow Neurological Institute, Phoenix, Arizona; ##Department of Neurological Surgery, Weill Cornell Medical College, New York Presbyterian Hospital, New York, New York; ***Department of Neurological Surgery, University of Washington School of Medicine, Seattle, Washington; ‡‡‡Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, Tennessee; §§§Inova Neuroscience Institute, Falls Church, Virginia

Portions of this work were presented in abstract form and poster form at the 2013 Congress of Neurological Surgeons Annual Meeting, in San Francisco, California.

†Deceased.

Correspondence:
Adam M. Sonabend, MD,
Department of Neurological Surgery,
Columbia University Medical Center,
710 W 168th St, Room 426,
New York, NY 10032.
E-mail: as3537@cumc.columbia.edu

Received, September 29, 2015.
Accepted, May 26, 2016.
Published Online, January 6, 2017.

Defining Glioblastoma Resectability Through the Wisdom of the Crowd: A Proof-of-Principle Study

BACKGROUND: Extent of resection (EOR) correlates with glioblastoma outcomes. Resectability and EOR depend on anatomical, clinical, and surgeon factors. Resectability likely influences outcome in and of itself, but an accurate measurement of resectability remains elusive. An understanding of resectability and the factors that influence it may provide a means to control a confounder in clinical trials and provide reference for decision making.

OBJECTIVE: To provide proof of concept of the use of the collective wisdom of experienced brain tumor surgeons in assessing glioblastoma resectability.

METHODS: We surveyed 13 academic tumor neurosurgeons nationwide to assess the resectability of newly diagnosed glioblastoma. Participants reviewed 20 cases, including digital imaging and communications in medicine-formatted pre- and postoperative magnetic resonance images and clinical vignettes. The selected cases involved a variety of anatomical locations and a range of EOR. Participants were asked about surgical goal, eg, gross total resection, subtotal resection (STR), or biopsy, and rationale for their decision. We calculated a "resectability index" for each lesion by pooling responses from all 13 surgeons.

RESULTS: Neurosurgeons' individual surgical goals varied significantly ($P = .015$), but the resectability index calculated from the surgeons' pooled responses was strongly correlated with the percentage of contrast-enhancing residual tumor ($R = 0.817$, $P < .001$). The collective STR goal predicted intraoperative decision of intentional STR documented on operative notes ($P < .01$) and nonresectable residual ($P < .01$), but not resectable residual.

CONCLUSION: In this pilot study, we demonstrate the feasibility of measuring the resectability of glioblastoma through crowdsourcing. This tool could be used to quantify resectability, a potential confounder in neuro-oncology clinical trials.

KEY WORDS: Resectability, Extent of resection, EOR, Glioblastoma, Glioma, Crowdsourcing

Neurosurgery 80:590–601, 2017 DOI:10.1227/NEU.0000000000001374 www.neurosurgery-online.com

Surgery plays multiple roles in glioblastoma (GBM) management, including obtaining tissue for pathological diagnosis and molecular studies, relieving neurological symptoms secondary to mass effect, and reducing steroid dependence. A series of studies have also shown a correlation between extent of resection (EOR) of gadolinium-enhancing tumor and survival, suggesting that surgery plays a role in oncological control.[1-19] Unlike the evidence supporting the efficacy of radiation therapy,[20] temozolomide,[21,22] bevacizumab,[23,24] and carmustine wafer implant,[25,26] studies evaluating the role of resection for oncological control are confounded by patient- and surgeon-related variables that influence EOR. EOR is determined by anatomical factors that affect whether a lesion can be resected in its entirety, ie, resectability. The resectability of a lesion likely influences surgeons' goals, such as whether to pursue gross total resection (GTR), subtotal resection (STR), or whether a patient should only undergo a biopsy (Bx). However, variation in providers' surgical goals remains poorly understood.

ABBREVIATIONS: **AI,** aggressiveness index; **ANOVA,** analysis of variance; **Bx,** biopsy; **DICOM,** digital imaging and communications in medicine; **EOR,** extent of resection; **GBM,** glioblastoma; **GTR,** gross total resection; **KPS,** Karnofsky Performance Scale; **MSM,** motor-speech-middle cerebral artery; **NIH score,** National Institutes of Health Recurrent GBM Score; **RI,** resectability index; **STR,** subtotal resection

Attempts to define the factors that determine resectability have been incorporated into clinical scales that help predict surgical morbidity[27] and patient outcomes[28] following GBM resection. Nevertheless, 1 modern study demonstrated only 37% agreement between 2 academic tumor surgeons on preoperative assessment of whether a given GBM was amenable to GTR.[29] Moreover, even when a tumor was determined to be amenable to GTR, this was only achieved in 25% of cases.[29] The influence of clinical judgment, intraoperative monitoring expertise, and tumor visualization techniques (eg, neuronavigation, intraoperative ultrasound, 5-aminolevulinic acid, intraoperative magnetic resonance imaging [MRI]) remain poorly understood, but are critical to the evaluation of EOR data.

In this proof-of-principle study, we investigated the reproducibility of resectability assessment for GBM. To do this, we used a crowdsourcing approach. Crowdsourcing refers to the recruitment of a large community via the Internet to solve a complex problem by taking advantage of the so-called "collective intelligence" or "wisdom of the crowd." This approach is widely used in commercial and manufacturing sectors and is emerging as an innovative tool for health care.[30] A series of studies are exploring the use of crowdsourcing by pooling multiple expert opinions on diagnostic dilemmas.[31-35] In neurosurgery, for instance, a crowdsourcing approach has been used to define the anatomical criteria leading to equipoise for anterior vs posterior cervical spine instrumentation and fusion.[36] Nevertheless, the crowdsourcing approach does not yet play a significant role in peer-to-peer collaboration for clinical practice.

Defining GBM resectability is challenging. In this study, we used crowdsourcing as a means of determining resectability with a highly selected patient data set designed to test anatomically complex scenarios. We focused on clinical judgment regarding a priori surgical goals as well as the presence and resectability of residual tumor.

METHODS

Clinical and Radiographic Data

Clinical and radiographic data were gathered from newly diagnosed GBM patients who underwent surgery at Columbia University Medical Center/New York Presbyterian (years 2000-2012). An initial set of 99 patients was identified with adequate MRI studies and clinical information. From this group, we selected 20 patients who are representative of lesions involving a variety of anatomical locations and with a wide range of resection outcomes. Patients were selected so that the data set would include a wide range of patients with regard to their tumor resectability. To do this, we purposefully included 2 patients who underwent Bx and 8 patients who underwent a craniotomy with STR in which the operating surgeon intentionally left residual disease deemed unsafe for further resection documented on the operative notes. The remaining 10 patients underwent a craniotomy for which GTR was thought to have been achieved according to operative notes (Table 1). All patients selected had a preoperative and postoperative contrast MRI performed within 48 hours of surgery, with the exception of 2 patients who underwent Bx. All clinical and radiographic data were deidentified before inclusion in the study as established by Health Insurance Portability and Accountability Act regulations and the institutional review board protocol at Columbia University. Volumetric assessment of the extent of resection of enhancing tumor was performed with image analysis tools MATLAB and automatic thresholding as previously described.[37] The cases were scored according to the National Institutes of Health Recurrent GBM Score (NIH score) and its motor-speech-middle cerebral artery (MSM) component in isolation.[28,38] In brief, the NIH score predicts survival following resection for recurrent GBM and takes into account a Karnofsky Performance Scale (KPS) score $\leq$80, tumor volume $\geq$50 cm^3, and the MSM score. The MSM score was developed to assess the resectability of low-grade gliomas. A higher score predicts decreased extent of resection, progression-free survival, and overall survival. Points are given for involvement of areas directly adjacent to the M1 and/or M2 segments of the middle cerebral artery[28] as well as eloquent regions including the sensorimotor strip, dominant superior temporal, inferior frontal and inferior parietal areas, basal ganglia, internal capsule, thalamus, and calcarine visual cortex.[38]

The Glioblastoma Resectability Survey

Three lines of questioning were associated with each patient's MRI (Figure 1). First, participants were asked to review digital imaging and communications in medicine (DICOM) images of preoperative T1 MRIs with and without contrast without any clinical information, then they were asked to select a surgical goal of GTR, STR, or Bx (anatomical assessment). If STR was selected, the surgeon was asked why the lesion was not amenable to GTR, ie, involvement of critical neural or nonneural structures. Subsequently, the software provided a clinical vignette including age, sex, KPS,[39] pertinent neurological findings, presenting symptoms, and comorbidities alongside the same DICOM images, and surgeons were asked the same questions regarding the surgical goal (clinical assessment). Finally, participants were asked to review postoperative T1 MRIs with and without contrast and assess whether there was residual tumor present. If the surgeon indicated that there was residual tumor, he or she was then asked whether the residual lesion was amenable to further resection. Participants who stated that the residual tumor was not resectable were asked to select the reason (involvement of critical neural structures, critical nonneural structures, or the presence of separate lesions).

Neurosurgeon Participants

An email invitation to participate was sent to 38 candidates who were faculty affiliated with an Accreditation Council for Graduate Medical Education-accredited neurosurgery residency program located at an academic medical center in the United States. All candidates invited had clinical practices that focus on brain tumors. To participate, surgeons were required to respond to the invitation, sign a consent form, and have an Apple computer running OS 10.7 or later with an Internet connection.

Software Platform and Data Collection

Dr William Butler designed an interactive software package that we used to survey surgeons. The software presents clinical vignettes, interfaces with OsiriX v.4.1.2 32-bit (http://www.osirix-viewer.com) to display DICOM images, and collects multiple-choice answers. By default, the software simultaneously displayed linked T1 MRIs with and without contrast for a given patient. The software, OsiriX, and the cases were contained on 64 GB USB drives. Every participating

TABLE 1. Patient Clinical Data and Clinical Vignettes[a,b]

Patient No.	Sex	Age	KPS	Clinical Vignette	Tumor Location	Surgery	Resection Stopped for Safety	Preop Contrast-Enhancing Volume, mm^3	Postop Contrast-Enhancing Residual, mm^3	Residual, %	Resectability Index	MSM Score	Variance in Surgeons' Goals
1	M	52	90	Word-finding difficulties; mild R hemiparesis, R toe upgoing, impaired R finger-nose-finger and heel-to-shin maneuvers	L. frontal	Resection	No	11 102	4570	41.1	0.79	1	0.47
2	F	49	80	Visual auras, vertigo, R homonymous hemianopsia	L. parietal	Resection	No	12 443	1966	15.8	0.25	1	0.39
3	F	56	60	Aphasia, confusion, syncopal episodes, dense R hemiparesis	R. frontal	Resection	Yes	87 051	30 688	35.2	1.08	3	0.58
4	M	46	80	L. inferior visual field defect, mild L hemiparesis, L sensory neglect	L. frontal	Resection	No	35025	1422	4	0.00	1	0
5	M	56	70	Severe headaches, personality changes, L focal seizures, mild L weakness, flat affect, cognitive impairment, mild L upper extremity weakness, L toe upgoing	R. frontal	Resection	Yes	121 498	6369	5.2	0.85	1	0.47
6	F	72	90	Progressive memory deficits, word-finding difficulty	L. temporal	Resection	Yes	47 871	11 347	23.7	0.43	3	0.44
7	M	62	80	New-onset seizures, memory deficits, word-finding difficulty, slurred speech, R facial weakness, mild R lower extremity weakness, unable to tandem walk	L. insular	Resection	Yes	16 852	15 099	89.6	1.23	4	0.69
8	M	60	90	Severe headaches, personality changes, cognitive impairment	R. frontal	Resection	No	39 824	519	1.3	0.21	0	0.36
9	F	74	80	Behavior changes, abulia, confusion, unable to tandem walk	L. frontal	Resection	No	39 237	5742	14.63	0.15	0	0.14
10	M	58	80	Mild L hemiparesis, decreased L hand fine-motor control	R. temporoin-sular, b. ganglia	Resection	Yes	69 179	10 350	14.96	0.92	2	0.18

TABLE 1. *Continued*

Patient No.	Sex	Age	KPS	Clinical Vignette	Tumor Location	Surgery	Resection Stopped for Safety	Preop Contrast-Enhancing Volume, mm^3	Postop Contrast-Enhancing Residual, mm^3	Residual, %	Resectability Index	MSM Score	Variance in Surgeons' Goals
11	F	77	70	Headache, receptive and expressive aphasia, cognitive impairment, paucity of speech, reverts to native language, unable to follow commands	L. frontotemporal	Resection	Yes	61 550	12 392	20.1	0.08	1	0.08
12	M	74	80	Headaches, slurred speech, memory deficit, lethargy, L facial droop	R. frontal	Resection	No	49 223	1180	2.3	0.54	2	0.27
13	F	58	80	L visual field loss, confusion	R. temporal	Resection	No	63 751	7086	11.1	0.23	2	0.19
14	M	76	80	Cognitive impairment, L homonymous hemianopsia, unstable gait, unable to tandem walk	R. parietal	Resection	No	31 670	9740	30.7	0.23	0	0.19
15	F	73	80	Headaches, confusion, mild L-sided weakness	R. frontal	Resection	Yes	77 394	15 746	20.3	0.54	0	0.27
16	M	61	90	Headaches, confusion, cognitive impairment	R. frontal	Resection	No	41 249	1785	4.3	0.62	1	0.59
17	F	64	80	Headaches, memory deficit, L homonymous hemianopsia	R. occipitotemporal	Resection	No	52 771	698	1.3	0.00	0	0
18	M	44	70	Increasing frequency of baseline seizures, new residual R-sided weakness, receptive and expressive aphasia	L. temporal	Resection	Yes	30 849	5182	16.7	0.75	1	0.75
19	F	18	80	Fatigue, malaise, headache, episodic emesis, diplopia, L sixth cranial nerve palsy, papilledema	R. frontal	Biopsy	No	12331	N/A	N/A	1.92	2	0.08
20	M	62	80	Severe fatigue, decreased visual acuity bilaterally, L upper extremity numbness	R. thalamic	Biopsy	No	30 054	N/A	N/A	1.54	2	0.44

[a]Bx, biopsy; GTR, gross total resection; KPS, Karnofsky performance status; L, left; postop, postoperative; preop, preoperative; R, right; STR, subtotal resection.
[b]Clinical data and clinical vignettes for the 20 patients whose images were included in the survey. Table also includes the values for the variance for surgical goal across all the surgeons for every patient. Resectability index (RI) was calculated by assigning quantitative values to surgeons' responses (GTR = 0, STR = 1, Bx = 2) and averaging the results for each patient. RI scores range from 0 to 2, with 0 being the most resectable and 2 being the least resectable.

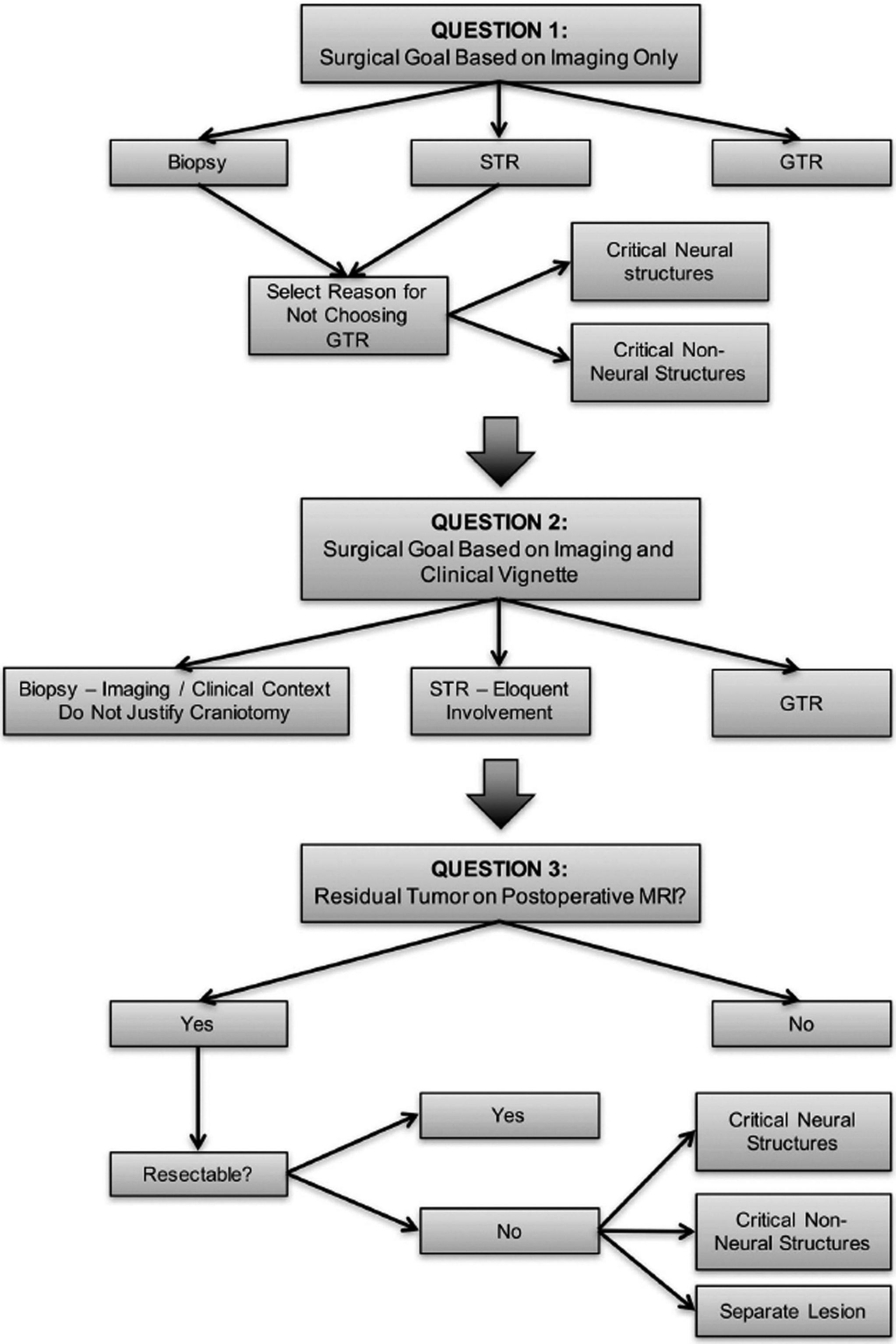

FIGURE 1. *Flowcharts showing the sequence of questions associated with each set of MRIs. GTR, gross total resection; STR, subtotal resection.*

neurosurgeon was assigned a USB drive containing the study material and an anonymous ID to log in to the software. Deidentified answers were automatically sent online in real time to a centralized database.

Statistical Analysis

Intersurgeon agreement within each question was determined using Fleiss Kappa and the intraclass correlation. Statistical analyses were performed with Microsoft Excel 2011, StatPlus (LE), Prism 6.0b, R, and MATLAB. P values of <.05 were considered significant.

In an attempt to quantify the aggressiveness of surgical decision making, we created a metric that we refer to as aggressiveness index (AI). A surgeon's AI is the average of his or her answer to the first question in the survey, with quantitative values assigned to each possible response (GTR = 0, STR = 1, and Bx = 2). Averages were calculated for each surgeon individually and for the group as a whole. The AI ranges from 0, being the most aggressive, to 2, being the least aggressive. When surgeons analyze the same patient cohort, the AI provides a summary view of their likelihood to pursue aggressive resection relative the mean of the cohort. Analysis of variance (ANOVA) was used to analyze differences in AI and was applied as follows: the quantitative values of each surgeon's answers for surgical goal for each patient were averaged, and the variance between these was calculated. Then, ANOVA was applied to these results, with significance reflecting differences in aggressiveness (as measured by AI) that are not likely the result of random variation. We used $P < .05$ as the threshold for significance.

To quantify resectability, we calculated a metric that we refer to as the "resectability index" (RI). As with the AI, responses were assigned quantitative values as follows: GTR = 0, STR = 1, and Bx = 2. RI scores range from 0 to 2, with 0 being the most resectable and 2 being the least resectable. We also measured the level of controversy among surgeons by calculating the variance in surgeons' responses for a given patient, using GTR = 0, STR = 1, and Bx = 2 for the variance calculation.

A biased association between surgical goal and surgical outcome could result from comparing a given surgeon's assessment of a lesion's resectability both preoperatively and postoperatively, because the surgeon's assessment is likely to stay consistent as opposed to providing 2 independent assessments (endogeneity bias). We devised a method to eliminate this bias while comparing pre- and postoperative assessments. We formed 2 groups consisting of 6 surgeons each, with no surgeon belonging to both groups. We used 1 group's responses as a set of preoperative assessments, and the other group's responses as a set of postoperative assessments. As described above, preoperative assessments were assigned quantitative values as follows: GTR = 0, STR = 1, and Bx = 2. Postoperative assessments were assigned quantitative values as follows: GTR with resectable residual was assigned a value of 0, and no resectable residual was assigned a value of 1. The average of the first and second group was calculated for each patient. This process was repeated for every combination of 6 surgeons in the pre- and postoperative assessment groups, resulting in 924 observations for each of the 18 patients who had pre- and postoperative images. For this analysis, a linear model was adjusted to correlate the postoperative assessment using preoperative assessment score averages as the independent variable.

RESULTS

Neurosurgeon Cohort

Thirteen neurosurgeons from 11 institutions met the inclusion criteria and completed the study with a wide geographical representation within the United States. Participants came from the following institutions: The Barrow Neurological Institute; Columbia University (2 participants); Cornell University; Dartmouth University; Vanderbilt University; Inova Neuroscience Institute, Falls Church, Virginia; Johns Hopkins University; Northwestern University; Massachusetts General Hospital; The University of California, San Diego; The University of Michigan (2 participants); University of Washington. These participants represent 37% of the neurosurgeons invited. All participants had an academic affiliation and a subspecialty in surgical neuro-oncology as reflected by their reputation, scope of practice, and peer-reviewed publications on brain tumors. Ninety-two percent of the participants were board certified in neurological surgery. The participants had an average of 18 years of clinical practice in neurosurgery (range of 4-33 years).

Resectability Criteria Among Surgeons

Variation among surgeons' responses showed fair interrater reliability for goal of surgery, ie, resection vs Bx ($\kappa = 0.286$; intraclass correlation = 0.435). As expected, there was more agreement about whether resection or Bx should be pursued (standard deviation, 0.3745) than whether GTR or STR should be the surgical goal (standard deviation, 0.4873). Surgeons' AI scores ranged from 0.21 to 1.0 (a lower AI value signifies that the surgeon tends to have a more aggressive surgical goal) (Figure 2). The mean AI score was 0.588 ± 0.274. Differences in aggressiveness among surgeons were statistically significant (ANOVA $P = .015$).

Resectability Among Patients

The majority of surgeons agreed on the surgical goal for most patients. Indeed, for 80% of the patient cohort, greater than or equal to 80% of surgeons were concordant in their assessment of resection vs Bx as the surgical goal. The 4 most controversial patients were identified according to their variance and compose the 80th percentile of variance in our sample (Table 1). There was no difference in the number of structures affected between the controversial patients and noncontroversial patients (Table 2). A greater proportion of the structures involved in the controversial patients' lesions were left-sided than in the noncontroversial patients' lesions (14/17 vs 21/56; RR 5.07 [1.59-16.16], $P = .002$). Lesions' RI scores ranged from 0 to 1.9 (Table 1). The mean score was 0.617 ± 0.525. Differences in resectability among lesions were statistically significant (ANOVA $P < .001$). The variance of surgeons' responses for each patient varied from 0 (patients 4 and 17) to 0.75 (patient 18) (Table 1). Imaging from the most resectable, least resectable, and most controversial patient (highest variance) have been included (Figure 3).

The Power of Resectability Defined by Crowdsourcing

The crowdsourcing approach demonstrated a powerful ability to predict EOR. First, we found a correlation between the percentage of surgeons electing GTR as the surgical goal and

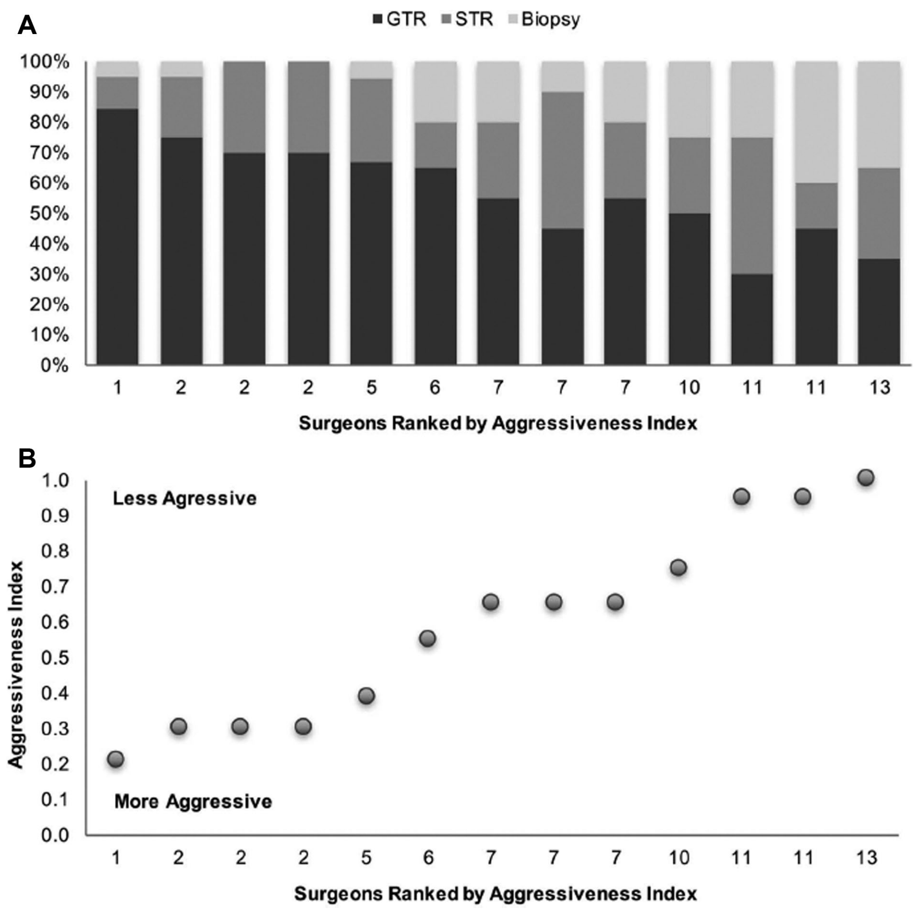

FIGURE 2. *The percentage of patients for whom each surgeon selected GTR, STR, and biopsy.* **A**, *surgeons are ranked by AI in descending order from left to right.* **B**, *a scatter plot representation of the AI for each individual surgeon, with the surgeons ranked by AI as above. A lower value on AI signifies a neurosurgeon who has more aggressive surgical goals, and a higher value represents more conservative surgical goals. AI, aggressiveness index; GTR, gross total resection; STR, subtotal resection.*

the percentage of surgeons who reported either that GTR was achieved or that the residual tumor was resectable ($R^2 = 0.77$, coefficient 0.7, 99% CI: 0.42-0.97, $P < .001$). The fraction of participants selecting STR as the surgical goal correlated with the intraoperative decision to pursue intentional STR, as documented in the operative notes ($P < .01$). Selecting STR as a surgical goal also predicted the postoperative assessment that the residual disease was not resectable ($P < .01$). The involvement of critical neural structures was the most commonly cited reason for selecting STR as the surgical goal and the most common reason for calling residual disease unresectable (Figure 4). Notably, surgeons varied with respect to their reason for pursuing STR. Goal of surgery based on radiographic information was not modified significantly by the addition of clinical information in the form of a vignette ($R^2 = 0.975$, coefficient 1.006, 99% CI: 0.89-1.1, $P < .0001$).

There is an inherent bias in comparing preoperative and postoperative assessments for the same surgeon. Surgeons provide both preoperative and postoperative assessments for each patient, and those assessments are likely consistent (endogeneity bias, see Methods). To control for this bias, we performed the following analysis. We formed all possible paired mutually exclusive groups of surgeons (924 groups). With these groups, we evaluated the correlation between preoperative surgical goal and the presence of resectable residual on postoperative MRI. We found that preoperative goal of GTR correlates with GTR or presence of resectable residual on postoperative assessment independent of endogeneity bias ($R = 0.444$, $P < .000$).

To further explore crowdsourcing's ability to predict EOR, we compared our RI (Methods) with volumetric data of enhancing residual tumor (Table 1). The RI had a moderate correlation with the volume of enhancing residual tumor ($R = 0.593$, slope

TABLE 2. Patient Lesion Anatomical Data[a]

Distinct Neural Structures	Total, n (%)		Controversial Patients, n (%)		Noncontroversial Patients, n (%)	
	Left	Right	Left	Right	Left	Right
Basal ganglia/internal capsule	4 (20)	5 (25)	2 (50)	1 (25)	2 (13)	3 (20)
Brainstem	0 (0)	1 (5)	0 (0)	1 (25)	0 (0)	0 (0)
Calcarine visual cortex	1 (5)	1 (5)	0 (0)	0 (0)	1 (7)	1 (7)
Corpus callosum	3 (15)	3 (15)	1 (25)	0 (0)	1 (7)	2 (13)
Sensorimotor strip	1 (5)	1 (5)	1 (25)	0 (0)	4 (27)	4 (27)
Inferior frontal	3 (15)	1 (5)	1 (25)	0 (0)	2 (13)	1 (7)
Inferior parietal	1 (5)	2 (10)	1 (25)	0 (0)	0 (0)	2 (13)
Superior temporal	5 (25)	3 (15)	1 (25)	0 (0)	2 (13)	1 (7)
Thalamus	1 (5)	3 (15)	0 (0)	0 (0)	1 (7)	1 (7)
Middle cerebral artery	3 (15)	1 (5)	1 (25)	0 (0)	0 (0)	4 (27)
Frontal	5 (25)	5 (25)	1 (25)	0 (0)	0 (0)	1 (7)
Temporal	6 (30)	7 (35)	2 (50)	0 (0)	3 (20)	3 (20)
Parietal	1 (5)	4 (20)	2 (50)	1 (25)	4 (27)	6 (40)
Occipital	1 (5)	1 (5)	1 (25)	1 (25)	0 (0)	1 (7)

[a]Anatomical structures involved in the lesions of the 20 patients in the survey. Percentages reflect the percent of patients within each category who have involvement of the structure in question. The controversial patients were defined as the 4 patients with the highest variance in responses of surgical goal. Patients whose lesions involve multiple structures are counted in multiple categories.

9842 mm^3/RI unit, intercept 3138, $P = .006$) and a strong correlation with the percentage of residual enhancing tumor ($R = 0.817$, slope 49.4% of tumor/RI unit, intercept -2.9, $P < .001$) (Figure 5).

We compared the ability of our crowdsourcing approach to predict volumetric EOR with other previously established scales that take into account radiographical variables related to eloquence and surgical morbidity for gliomas. Thus, we investigated whether the MSM score and the NIH score also predict EOR. The MSM score assesses the involvement of critical/eloquent brain structures on imaging, and the NIH score takes into account MSM plus other clinical variables.[28,38] Both the MSM score and the NIH score had considerably weaker correlations with both the volume and the percentage of residual enhancing lesion than the RI from the crowdsourcing approach (Figure 5, Table 3).

DISCUSSION

We present a pilot study using a novel instrument to assess surgical goal for GBM. We show that crowdsourcing is a useful approach for determining the resectability of these tumors. The systematic study of GBM resectability is complex, because it is influenced by surgeon- and patient-related factors, including anatomical and clinical variables that may not be captured by a rigid scoring system. Assessments from experienced surgeons likely take such variables into account.

One of the motivations for performing this study is to investigate the predictive power of neurosurgeons for determining the outcome of a tumor surgery. Patients make decisions based on a surgeon's estimation of risks and benefits, including the expected EOR. However, the certainty of this individual EOR estimation is unknown. Studies such as ours aim to provide a reference for the accuracy of such predictions. It is important to stress that the accuracy of the preoperative estimation of EOR is determined not only by the lesion's resectability, but also by the surgeon's ability, which can vary across surgeons. Our study only investigates the former of these 2 determinants of EOR. Yet, by measuring and controlling for resectability on large data sets, future studies could tease apart other variables that might influence EOR, such as a surgeon's ability.

This study identified interesting differences in surgical goals. Using a highly curated set of 20 cases for which controversy is expected, we encountered significant differences in surgical aggressiveness among 13 academic tumor neurosurgeons, but less variation than previously reported in the literature. It is noteworthy that the differences in surgical goals included important discrepancies not only on how much should be resected, but also on whether a biopsy should be done instead of a craniotomy for resection. We also found significant variation in the reason for pursuing STR in cases for which STR was the goal. Such differences in judgment may also account for why some surgeons selected STR for a given patient, but others selected GTR. Notably, our degree of interobserver agreement was high compared with previous studies. Orringer et al[29] showed only 37% agreement between 2 academic tumor surgeons on preoperative assessment of whether a GBM was amenable to GTR, and being assessed as amenable to GTR was only predictive of GTR in 25% of cases. It is likely that the higher consistency in responses seen in our study derives from a larger number

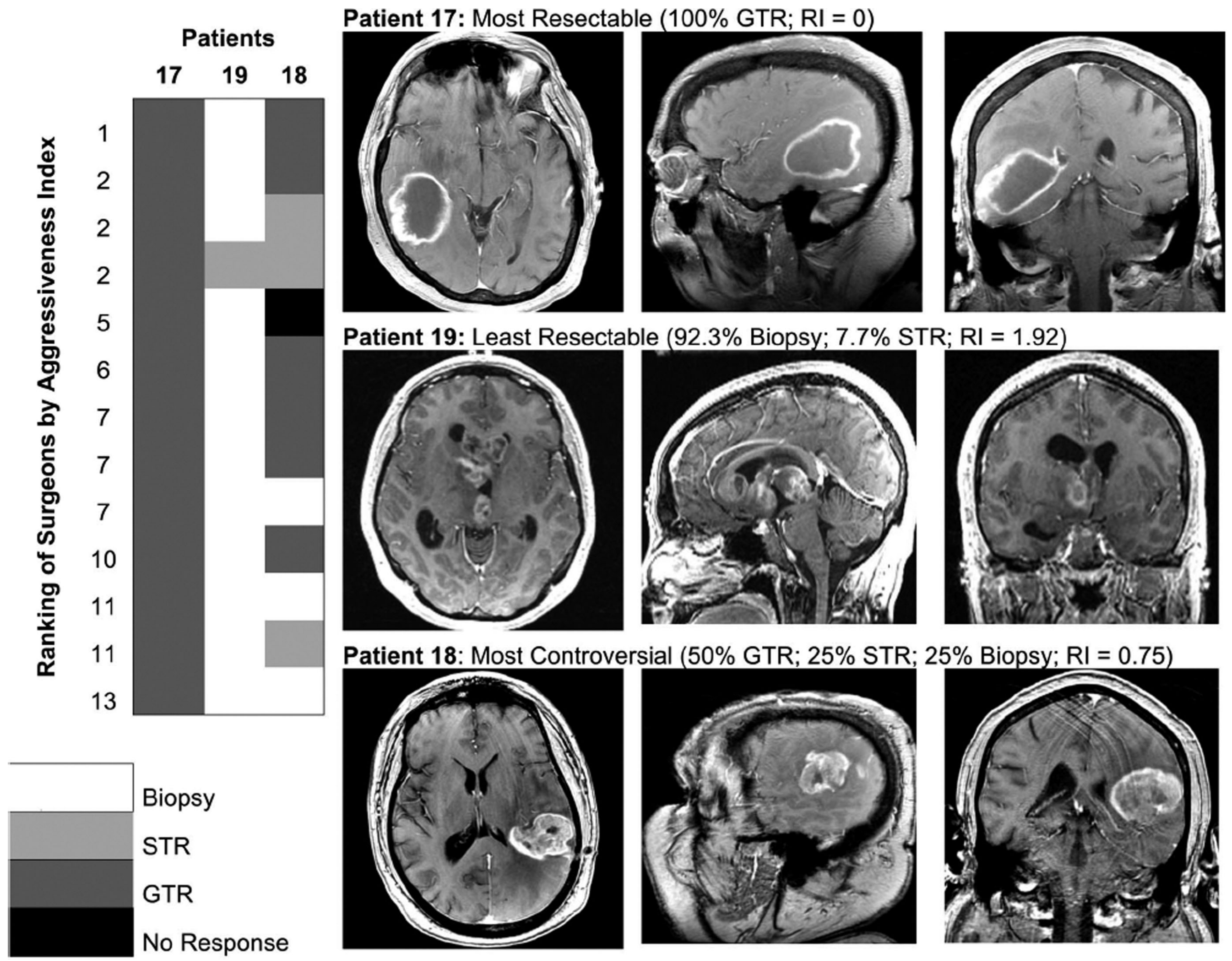

FIGURE 3. *MRI images in 3 planes for the most resectable (patient 17), least resectable (patient 19), and most controversial patient (patient 18), as determined by level of variance. See Table 1 for variance of responses for surgical goals. A heat map illustrating each surgeon's management plan for these 3 patients is also included, with white representing biopsy, light gray representing subtotal resection, dark gray representing gross total resection, and black representing no response. Surgeons are ranked by AI from most aggressive (***top***) to least aggressive (***bottom***). AI, aggressiveness index; GTR, gross total resection; RI, resectability index; STR, subtotal resection.*

of surgeons queried, supporting the role of crowdsourcing to determine resectability.

Crowdsourcing is a powerful approach to investigate resectability, yet it is time consuming and logistically challenging. We investigated the value of established scoring systems as an alternative to our approach. Two existing clinical/imaging-based scoring systems applied to our patients showed some correlation with the volume of enhancing residual on postoperative MRI. Notably, these scoring systems were not designed to predict resectability per se and were not designed or validated for newly diagnosed GBM. However, we chose these scales as a comparison, because these are reproducible scales that take into account anatomical/imaging variables that are presumed to be associated with eloquence, surgical morbidity, and outcome. The MSM score quantifies involvement of 3 critical brain areas demonstrated to be associated with decreased progression-free survival and decreased overall survival.[38] On the other hand, the NIH score for recurrent GBM incorporates the MSM score, as well as KPS and tumor volume, to predict postoperative survival following reoperation.[28] The correlation between both of these scores with the decision to pursue GTR demonstrates that the factors determining these scores affect surgeons' decision making. However, the relatively weak correlations of these scores with volume and percent of enhancing residual, compared with

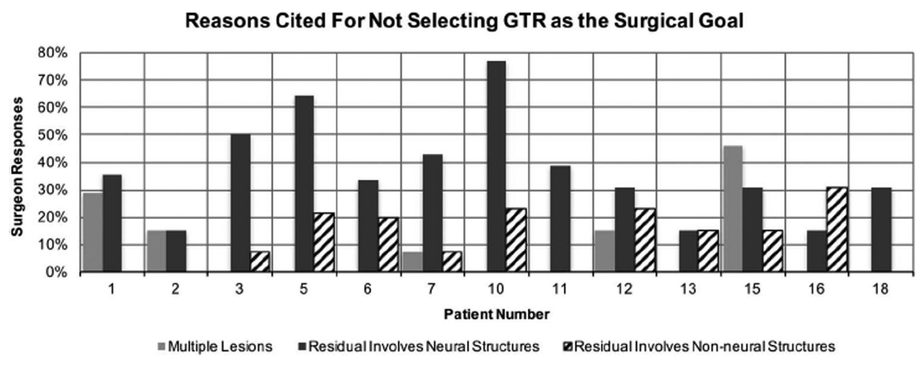

FIGURE 4. *Reasons cited for not selecting GTR as the surgical goal for residual disease on postoperative imaging. For each patient, bars depict the percentage of surgeons selecting a response on a multiple-choice menu presented for choosing the reason for not electing to pursue GTR. Patients for whom all surgeons reported that GTR was achieved, patients for whom all surgeons agreed that the residual disease was resectable, and patients who underwent a biopsy were excluded from the analysis. GTR, gross total resection.*

TABLE 3. Correlations Between Clinical Scales and Resection Outcome[a,b]

	Volume Residual				Percent Residual			
Independent Variable	***R***	**Slope**	**Slope 95% CI**	***P***	***R***	**Slope**	**Slope 95% CI**	***P***
NIH score	0.529	5219	1071-9369	0.017	0.233	8.4	−8.9 to 25.7	0.322
MSM score	0.497	3805	511-7100	0.026	0.515	14.4	2.5 to 26.3	0.020
Resectability index	0.593	9841	3230-16 454	0.006	0.817	49.4	32.1 to 66.7	<0.001

[a]MSM, motor-speech-middle cerebral artery component.

[b]The results of linear regressions relating either percentage of residual tumor or volume of residual enhancing tumor to resectability index (RI), MSM score, and NIH score. The slope for the volume regressions is mm^3 per unit on the scale in question. The slope for the percent residual regression is percent of residual tumor per unit on the scale question.

the strong correlation between the RI derived from surgeons' goals, suggest that the former capture only some of the factors contributing to surgeons' decisions.

Novel technologies to improve EOR for GBM, such as 5-aminolevulinic acid, are being evaluated in clinical trials, including studies for which being amenable to GTR is part of the inclusion criteria.[40-43] Our data suggest that there might be considerable heterogeneity in this inclusion criteria. However, without assessing the a priori resectability of the tumors, outcomes are likely to be confounded by uncontrolled anatomical and clinical factors influencing resectability, independent of the EOR. To resolve such issues, our approach might be used to stratify accordingly.

Recent studies have demonstrated that crowdsourcing can be successfully used in clinical medicine, most notably in pathology and ophthalmology.[31-35] Such novel methods could be used in neurosurgery. Furthermore, machine-learning tools, which have been developed for oncological prognostication, could be adapted to assist in neurosurgical decision making by pooling and analyzing crowdsourcing results.[44,45]

Limitations

There are several significant limitations of this study that should be considered. First, it was designed as a proof-of-concept study; therefore, both the number of surgical cases and neurosurgeon participants was small, limiting the generalization of these results. In many ways, this neurosurgeon cohort was rather homogeneous and may not reflect the perspectives of the general neurosurgical community. The true strength of crowdsourcing lies in the ability to pull the opinions and expertise of large groups. As we go on to further streamline the data acquisition process and improve the user interface, our aim is to distribute the survey on a large scale to a diverse set of participants. Additionally, this highly curated data set was chosen to be particularly controversial and, thus, may contribute to an increase in the variability among surgeons' responses.

Overall, systematic evaluation of GBM resectability might be critical for prognostication and evaluating outcomes in neuro-oncology trials. It is foreseeable that anatomical features that determine resectability could influence outcome independent of EOR. In fact, certain biological features, such as IDH1 mutation,

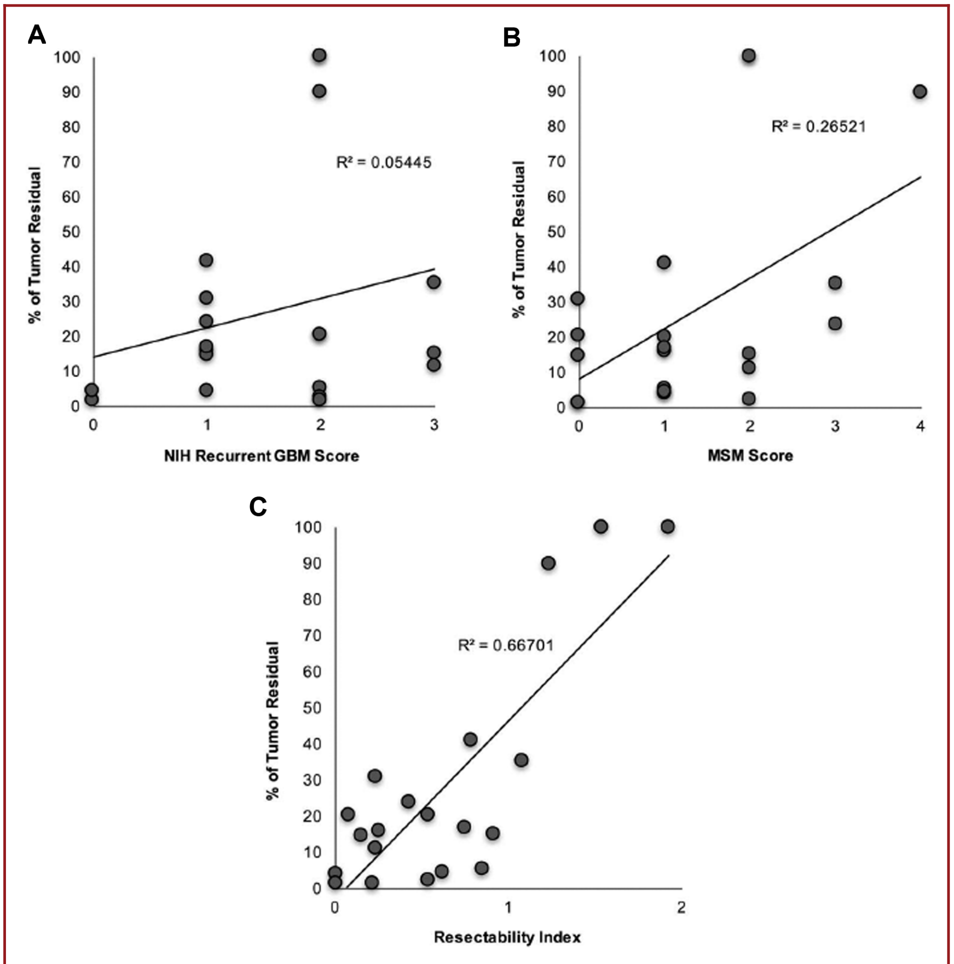

FIGURE 5. *Scatter plots with linear regressions relating the percentage of residual enhancing tumor to NIH score (slope = 8.4 [−8.9 to 25.7], R = 0.233, P = .322)* **A**; *MSM score (slope = 14.4 [2.5-26.3], R = 0.515, P = .020)* **B**; *and resectability index (RI) (slope = 49.4 [32.1-66.7], R = 0.817, P < .001)* **C**. *Slope is defined as percentage of tumor/RI unit. MSM, motor-speech-middle cerebral artery component.*

have been shown to be prognostic factors and to be associated with resectability.[46,47] This supports the possibility of resectability being a prognostic factor related to location and tumor biology that might be independent of the presence of residual disease. A larger study with additional surgeons and cases might allow exploration of these possibilities.

CONCLUSION

This pilot study demonstrates the feasibility of investigating resectability and surgical goal for GBM surgery, as well as resectability of residual disease through crowdsourcing. This tool could be used to quantify resectability, a potential confounder in neuro-oncology; however, additional studies are necessary to validate the utility of this approach.

Disclosures

This work was funded by the NIH Office of the Director, Award Number 1DP5OD021356-01, and a Diversity Recruitment Award by the Office of Senior Associate Provost for Faculty Diversity and Inclusion, Columbia University. The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Lacroix M, Abi-Said D, Fourney DR, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg.* 2001;95(2):190-198.
2. Sanai N, Polley MY, McDermott MW, Parsa AT, Berger MS An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg.* 2011;115(1):3-8.
3. Keles GE, Anderson B, Berger MS. The effect of extent of resection on time to tumor progression and survival in patients with glioblastoma multiforme of the cerebral hemisphere. *Surg Neurol.* 1999;52(4):371-379.

4. Lamborn KR, Chang SM, Prados MD. Prognostic factors for survival of patients with glioblastoma: recursive partitioning analysis. *Neuro Oncol.* 2004; 6(3):227-235.
5. Shibamoto Y, Yamashita J, Takahashi M, Yamasaki T, Kikuchi H, Abe M. Supratentorial malignant glioma: an analysis of radiation therapy in 178 cases. *Radiother Oncol.* 1990;18(1):9-17.
6. Simpson JR, Horton J, Scott C, et al. Influence of location and extent of surgical resection on survival of patients with glioblastoma multiforme: results of three consecutive Radiation Therapy Oncology Group (RTOG) clinical trials. *Int J Radiat Oncol Biol Phys.* 1993;26(2):239-244.
7. Jeremic B, Grujicic D, Antunovic V, Djuric L, Stojanovic M, Shibamoto Y. Influence of extent of surgery and tumor location on treatment outcome of patients with glioblastoma multiforme treated with combined modality approach. *J Neurooncol.* 1994;21(2):177-185.
8. Barker FG, Prados MD, Chang SM, et al. Radiation response and survival time in patients with glioblastoma multiforme. *J Neurosurg.* 1996;84(3):442-448.
9. Ushio Y, Kochi M, Hamada J, Kai Y, Nakamura H. Effect of surgical removal on survival and quality of life in patients with supratentorial glioblastoma. *Neurol Med Chir (Tokyo).* 2005;45(9):454-460; discussion 460-451.
10. Stark AM, Nabavi A, Mehdorn HM, Blomer U. Glioblastoma multiforme-report of 267 cases treated at a single institution. *Surg Neurol.* 2005;63(2):162-169; discussion 169.
11. Stummer W, Reulen HJ, Meinel T, et al. Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. *Neurosurgery.* 2008;62(3):564-576; discussion 564-576.
12. Sanai N, Berger MS. Glioma extent of resection and its impact on patient outcome. *Neurosurgery.* 2008;62(4):753-764; discussion 264-756.
13. Carrabba G, Fava E, Giussani C, et al. Cortical and subcortical motor mapping in rolandic and perirolandic glioma surgery: impact on postoperative morbidity and extent of resection. *J Neurosurg Sci.* 2007;51(2):45-51.
14. Woodward DE, Cook J, Tracqui P, Cruywagen GC, Murray JD, Alvord EC. A mathematical model of glioma growth: the effect of extent of surgical resection. *Cell Prolif.* 1996;29(6):269-288.
15. Iacoangeli M, Roselli R, Prezioso A, Scerrati M, Rossi GF. Staging of supratentorial hemispheric glioma using tumour extension, histopathological grade and extent of surgical resection. *Br J Surg.* 1993;80(9):1130-1133.
16. Winger MJ, Macdonald DR, Cairncross JG. Supratentorial anaplastic gliomas in adults. The prognostic importance of extent of resection and prior low-grade glioma. *J Neurosurg.* 1989;71(4):487-493.
17. Ammirati M, Vick N, Liao YL, Ciric I, Mikhael M. Effect of the extent of surgical resection on survival and quality of life in patients with supratentorial glioblastomas and anaplastic astrocytomas. *Neurosurgery.* 1987;21(2):201-206.
18. Ciric I, Ammirati M, Vick N, Mikhael M. Supratentorial gliomas: surgical considerations and immediate postoperative results. Gross total resection versus partial resection. *Neurosurgery.* 1987;21(1):21-26.
19. Kiwit JC, Floeth FW, Bock WJ. Survival in malignant glioma: analysis of prognostic factors with special regard to cytoreductive surgery. *Zentralbl Neurochir.* 1996;57(2):76-88.
20. Walker MD, Alexander E, Hunt WE, et al. Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. A cooperative clinical trial. *J Neurosurg.* 1978;49(3):333-343.
21. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005;352(10):987-996.
22. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* 2009;10(5):459-466.
23. Vredenburgh JJ, Desjardins A, Herndon JE, et al.. Bevacizumab plus irinotecan in recurrent glioblastoma multiforme. *J Clin Oncol.* 2007;25(30):4722-4729.
24. Cohen MH, Shen YL, Keegan P, Pazdur R. FDA drug approval summary: bevacizumab (Avastin) as treatment of recurrent glioblastoma multiforme. *Oncologist.* 2009;14(11):1131-1138.
25. Westphal M, Hilt DC, Bortey E, et al. A phase 3 trial of local chemotherapy with biodegradable carmustine (BCNU) wafers (Gliadel wafers) in patients with primary malignant glioma. *Neuro Oncol.* 2003;5(2):79-88.
26. Brem H, Ewend MG, Piantadosi S, Greenhoot J, Burger PC, Sisti M. The safety of interstitial chemotherapy with BCNU-loaded polymer followed by radiation therapy in the treatment of newly diagnosed malignant gliomas: phase I trial. *J Neurooncol.* 1995;26(2):111-123.
27. 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; discussion 1055-1046.
28. Park JK, Hodges T, Arko L, et al. Scale to predict survival after surgery for recurrent glioblastoma multiforme. *J Clin Oncol.* 2010;28(24):3838-3843.
29. 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.
30. Sims MH, Fagnano M, Halterman JS, Halterman MW. Provider impressions of the use of a mobile crowdsourcing app in medical practice. *Health Informatics J.* 2016;22(2):221-231.
31. Della Mea V, Maddalena E, Mizzaro S, Machin P, Beltrami CA. Preliminary results from a crowdsourcing experiment in immunohistochemistry. *Diagn Pathol.* 2014;9(suppl 1):S6.
32. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. *J Med Internet Res.* 2014;16(10):e233.
33. Irshad H, Montaser-Kouhsari L, Waltz G, et al. Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. *Pac Symp Biocomput.* 2015;20:294-305.
34. Mitry D, Peto T, Hayat S, et al. Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. *PLoS One.* 2015;10(2):e0117401.
35. Mitry D, Peto T, Hayat S, Morgan JE, Khaw KT, Foster PJ. Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk cohort on behalf of the UK Biobank Eye and Vision Consortium. *PLoS One.* 2013;8(8):e71154.
36. Ghogawala Z, Coumans JV, Benzel EC, Stabile LM, Barker FG. Ventral versus dorsal decompression for cervical spondylotic myelopathy: surgeons' assessment of eligibility for randomization in a proposed randomized controlled trial: results of a survey of the Cervical Spine Research Society. *Spine (Phila Pa 1976).* 2007;32(4):429-436.
37. Swanson KR, Chakraborty G, Wang CH, et al. Complementary but distinct roles for MRI and 18F-fluoromisonidazole PET in the assessment of human glioblastomas. *J Nucl Med.* 2009;50(1):36-44.
38. Chang EF, Smith JS, Chang SM, et al. Preoperative prognostic classification system for hemispheric low-grade gliomas in adults. *J Neurosurg.* 2008;109(5): 817-824.
39. Mor V, Laliberte L, Morris JN, Wiemann M. The Karnofsky Performance Status Scale. An examination of its reliability and validity in a research setting. *Cancer.* 1984;53(9):2002-2007.
40. Senft C, Bink A, Franz K, Vatter H, Gasser T, Seifert V. Intraoperative MRI guidance and extent of resection in glioma surgery: a randomised, controlled trial. *Lancet Oncol.* 2011;12(11):997-1003.
41. Senft C, Bink A, Heckelmann M, Gasser T, Seifert V. Glioma extent of resection and ultra-low-field iMRI: interim analysis of a prospective randomized trial. *Acta Neurochir Suppl.* 2011;109:49-53.
42. Senft C, Seifert V, Hermann E, Franz K, Gasser T. Usefulness of intraoperative ultra low-field magnetic resonance imaging in glioma surgery. *Neurosurgery.* 2008; 63(4 Suppl 2):257-266; discussion 266-257.
43. Hatiboglu MA, Weinberg JS, Suki D, et al. Impact of intraoperative high-field magnetic resonance imaging guidance on glioma surgery: a prospective volumetric analysis. *Neurosurgery.* 2009;64(6):1073-1081; discussion 1081.
44. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2015;13:8-17.
45. Enshaei A, Robson CN, Edmondson RJ. Artificial intelligence systems as prognostic and predictive tools in Ovarian Cancer. *Ann Surg Oncol.* 2015;22 (12):3970-3975.
46. Beiko J, Suki D, Hess KR, et al. IDH1 mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro Oncol.* 2014;16(1):81-91.
47. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *New Engl J Med.* 2009;360(8):765-773.