Balancing task sensitivity with reliability for multimodal language assessments

Alexander A. Aabedi, BS[1], **Sofia Kakaizada, BS**[1], **Jacob S. Young, MD**[1], **EunSeon Ahn, BS**[2], **Daniel H. Weissman, PhD**[2], **Mitchel S. Berger, MD**[1], **David Brang, PhD**[2], **Shawn L. Hervey-Jumper, MD**[1]

[1]Department of Neurological Surgery, University of California, San Francisco, California

[2]Department of Psychology, University of Michigan, Ann Arbor, Michigan

Abstract

OBJECTIVE—Intraoperative tasks for awake language mapping are typically selected based on the language tracts that will likely be encountered during tumor resection. However, diminished attention and arousal secondary to perioperative sedatives may reduce a task's usefulness for identifying eloquent cortex. For instance, accuracy in performing select language tasks may be high preoperatively but decline in the operating room. In the present study, the authors sought to identify language tasks that can be performed with high accuracy in both situational contexts so the neurosurgical team can be confident that speech errors committed during awake language mapping result from direct cortical stimulation to eloquent cortex, rather than from poor performance in general.

METHODS—We administered five language tasks to 44 patients: picture naming (PN), text reading (TR), auditory object naming (AN), repetition of 4-syllable words (4SYL), and production of syntactically intact sentences (SYNTAX). Performance was assessed using the 4-point scale of the quick aphasia battery 24 hours preoperatively and intraoperatively. We next determined whether or not accuracy on each task was higher preoperatively than intraoperatively. We also determined whether 1) intraoperative accuracy on a given task predicted intraoperative performance on the other tasks and 2) low preoperative accuracy on a task predicted a decrease in accuracy intraoperatively.

RESULTS—Relative to preoperative accuracy, intraoperative accuracy declined on PN (3.90 vs 3.82, p = 0.0001), 4SYL (3.96 vs 3.91, p = 0.0006), and SYNTAX (3.85 vs 3.67, p = 0.0001) but

not on TR (3.96 vs 3.94, p = 0.13) or AN (3.70 vs 3.58, p = 0.058). Intraoperative accuracy on PN and AN independently predicted intraoperative accuracy on the remaining language tasks (p < 0.001 and p < 0.01, respectively). Finally, low preoperative accuracy on SYNTAX predicted a decrease in accuracy on this task intraoperatively (R^2 = 0.36, p = 0.00002).

CONCLUSIONS—While TR lacks sensitivity in identifying language deficits at baseline, accuracy on TR is stable across testing settings. Baseline accuracy on the other four of our five language tasks was not predictive of intraoperative performance, signifying the need to repeat language tests prior to stimulation mapping to confirm reliability.

Keywords

awake craniotomy; language mapping; language task reliability; surgical technique

---

In the management of patients with infiltrative tumors in the dominant hemisphere, awake craniotomy with intraoperative language mapping is the most robust method of identifying the margins of safe lesion resection.[1–4] By maximizing tumor cytoreduction while preserving functional pathways involved in the generation and comprehension of speech, neurosurgeons can offer their patients both improved survival and preservation of quality of life.[5–10]

The diversity of available language tasks presents a salient opportunity to tailor intraoperative assessments to each patient's unique anatomical and functional connectivity within the proposed resection plane.[11–13] However, the lack of standardization, clinical guidelines, and evidence for each task's reliability in facilitating accurate mapping threatens their theoretical utility and must be balanced against the desire to remove as much tumor as possible.[14] For instance, administration of a task that is completed with low accuracy at baseline may lead to excessive false positives during stimulation mapping. Low task specificity carries the risk of miscategorization of noneloquent cortex as eloquent, which can result in erroneous preservation of "nonfunctional" regions and lead to an overly conservative resection.[15]

To mitigate this risk, various authors have recommended that patients receive language assessments 1 day prior to surgery and that only tasks completed with high baseline accuracy be included in the intraoperative assessments.[1,2] However, we have previously shown that nonlanguage task accuracy decreases in the operating room in a task-specific manner even after controlling for baseline performance.[16] This reduction may reflect fluctuations of attention and arousal that are induced by the lingering effects of sedatives, anxiolytics, and opioids, which are given to patients preoperatively to facilitate safe and comfortable performance of the craniotomy.

It remains unclear, though, whether certain language tasks are more susceptible to intraoperative declines in accuracy (and thus potentially less useful for language mapping). It also remains unclear which speech modalities provide the highest and most stable performance over time. Resolving these ambiguities would enable neurosurgical teams to conclude with greater certainty that speech errors during mapping result from electrical stimulation rather than perioperative anesthetics. Thus, in the present study, we employed

preoperative and intraoperative multimodal language assessments to investigate which of the several language tasks that are commonly used during awake mapping are best suited for identifying intraoperative language deficits.

Methods

We analyzed 44 patients from a single-institution, pro-spectively maintained registry of patients with dominant-hemisphere tumors presenting for resection with awake language mapping.[16,17] Language dominance was confirmed via preoperative magnetoencephalography. All patients received standard-of-care treatment by participating in this study, which was approved by our institutional review board.

According to established protocol, each patient underwent a language battery at baseline (1 day preoperatively) consisting of five tasks and 221 total trials.[17] These tasks included picture naming (PN), text reading (TR), auditory object naming (AN), repetition of 4-syllable words (4SYL), and description of a scene with correct syntax (SYNTAX). All tasks were delivered on a computer, which randomly ordered the tasks and trials for each assessment.

In the operating room, each patient received either propofol or dexmedetomidine according to the anesthetist's preference and subsequently underwent craniotomy.[18] After adequate exposure of the cortex was achieved, all anesthetics were stopped and the patient was given a minimum of 20 minutes to awaken. Patients were permitted to fluctuate within 2.5 SDs (derived from a pooled sample of 75 patients from our prospective registry) of their preoperative score on a vigilance task in order to proceed with language testing, given its established association with intraoperative language decline.[16] Each patient subsequently underwent the same language assessments as the previous day, again in a randomized fashion. To minimize operating room noise, the following measures were rigorously enforced during cognitive testing: 1) all staff and assistants were notified to minimize verbal communication, 2) telephones and alarms were muted and only allowed to make visual notifications, and 3) surgical suction, along with all other nonessential machinery, was turned off. All trials were scored by a qualified speech-language pathologist or a trained clinical research coordinator using the standardized quick aphasia battery (QAB).[19] After the intraoperative assessments were completed, the patients underwent stimulation mapping, maximal safe tumor resection, and closure, according to previously described protocols.[1,3]

Statistical Analysis

All statistical analyses were conducted using R (version 3.6.2). Specifically, the analyses were conducted by an author who was 1) not involved in scoring the assessments and 2) initially blinded to each subject's identity and clinical course. No patients were excluded from the analysis. However, 6 patients could not complete the full range of testing due to significant anxiety or refusal to continue. Each of these 6 patients had at least 8 complete tasks (i.e., the full extent of prespecified trial stimuli) available for analysis (Supplemental Table 1).

We conducted the statistical analyses as follows. Means were reported with interquartile ranges. Comparisons between preoperative and intraoperative task performance were computed using the Wilcoxon signed-rank test for paired, nonparametric samples. Between-task comparisons were made with one-way ANOVA. Relationships between continuous variables were assessed by fitting univariate generalized linear models. An alpha level of 0.05 was used to denote statistical significance. Finally, corrections for multiple comparisons were made using the false discovery rate. Our comparisons were 80% powered to identify a medium effect size (Cohen's d of 0.60) using a significance level of 0.05.

Results

Patient demographics and clinical characteristics are summarized in Table 1. Of the 44 patients we tested, 15 were female (34%). The mean age was 50.5 years (range 19–81 years). Tumors were predominantly in the left hemisphere (93%) and located in the frontal (36%) and temporal (32%) lobes. Six patients presented for reoperation for recurrence, while the remaining 38 underwent primary resection. Nineteen patients (43%) had WHO grade IV tumors on pathologic examination, 12 (27%) had WHO grade III tumors, 10 (23%) had WHO grade II tumors, 2 (7%) had metastases, and 1 (3%) had a cavernous malformation.

Between-Task Comparisons

Mean QAB scores on each of the five language tasks in the preoperative and intraoperative settings are summarized in Fig. 1. In the preoperative setting, one-way ANOVA revealed a significant main effect: AN task accuracy (mean QAB score = 3.70) was significantly lower than the accuracy on four other language tasks (PN, TR, 4SYL, and SYNTAX; $p = 0.000003$). In post hoc analyses, there were no statistically significant differences in mean QAB scores on any other tasks.

In the intraoperative setting, patients similarly performed the worst on AN (mean QAB score = 3.58). However, SYNTAX task accuracy decreased in the operating room such that the difference in QAB scores between AN and SYNTAX was no longer statistically significant (3.58 vs 3.67, $p = 0.83$). Accuracy on AN was significantly lower than on PN (3.58 vs 3.82, $p = 0.04$), TR (3.58 vs 3.94, $p = 0.002$), and 4SYL (3.58 vs 3.91, $p = 0.0002$). SYNTAX task accuracy was lower than only TR (3.67 vs 3.94, $p = 0.01$) and 4SYL (3.67 vs 3.91, $p = 0.048$).

Within-Task Comparisons

Paired comparisons between preoperative and intraoperative testing sessions for each of the five language tasks are summarized in Fig. 2. Compared with preoperative testing, there were statistically significant declines of intraoperative accuracy on PN (3.90 vs 3.82, $p = 0.0001$), 4SYL (3.96 vs 3.91, $p = 0.0006$), and SYNTAX (3.85 vs 3.67, $p = 0.0001$). There was also a trend toward lower accuracy on AN, which did not reach statistical significance (3.70 vs 3.58, $p = 0.058$). TR accuracy remained stable across testing environments (3.96 vs 3.94, $p = 0.13$).

Dumbbell plots illustrating the change in performance between preoperative and intraoperative testing for each patient across all five language tasks are displayed in Fig. 3. Certain patients (i.e., patients 19 and 25) experienced universal declines in language task accuracy compared with preoperative performance, while others (i.e., patients 32 and 33) had relatively modest declines that were limited to one or two language tasks.

Predicting Changes in Language Task Accuracy

To determine whether intraoperative accuracy on one language task predicted intraoperative accuracy on the remaining language tasks, we ran univariate comparisons between each task and the four other tasks (Table 2). Low intraoperative task accuracy on PN and AN was an independent predictor of low accuracy on all other tasks. Thus, if a patient performed poorly on either PN or AN intraoperatively, they were more likely to perform poorly on the remaining language tasks in the operating room. TR, 4SYL, and SYNTAX were predictive of performance on some, but not all, of the remaining language tasks.

Finally, in order to predict whether select patients are at risk of intraoperative declines, we performed linear regressions between preoperative task accuracy and the percent change in task accuracy between settings (Fig. 4). While patients with high preoperative SYNTAX scores had intact performance in the operating room, low pre-operative SYNTAX scores were associated with subsequent intraoperative declines in accuracy ($R^2 = 0.36$, p = 0.00002). There were no significant associations between preoperative scores on PN, TR, AN, and 4SYL and subsequent intraoperative changes from baseline on those tasks.

Discussion

Language tasks for stimulation mapping are selected based on a patient's underlying brain anatomy and presumed neural networks lying within and around the intended resection plane.[20,21] However, the stability of many common language tasks across testing sessions is not thoroughly established in the literature and, consequently, not typically included in decision-making paradigms for language task selection. Therefore, the goal of the present study was to determine whether or not accuracy on five language tasks that are frequently used for awake language mapping declines during intraoperative testing. More specifically, we hoped to identify tasks that are robust to the reductions of arousal and attention from perioperative anesthetics. Using such tasks would allow neurosurgical teams to conclude with greater certainty that speech errors made during mapping result from electrical stimulation rather than from effects of perioperative anesthetics.

The most commonly used task in awake language mapping is visual object naming (PN in this study). This task allows the neurosurgical team to determine whether a stimulated region of interest is involved in visuoperceptual processing, speech production, and/or semantic integration.[14] Given the diffuse functional and anatomical circuits that are required to perform PN, this task provides excellent sensitivity for identifying eloquent regions of cortex that must be preserved to prevent postoperative language deficits.[22] However, patients may perform this task inaccurately in the intraoperative context even before mapping begins. In these situations, neurosurgeons will often abort awake language mapping and revert to asleep image-guided resection if dysphasia becomes significant.[23] Thus, it is important to

determine whether other tasks may be substituted for PN in such situations and, more generally, which tasks allow the most stable performance over time.

Of the five language tasks used in this study, TR demonstrated the most stability across testing environments with restricted distributions around the "ceiling" (i.e., the highest possible QAB score of 4). PN accuracy (along with 4SYL and SYNTAX), on the other hand, declined significantly in the operating room. This result demonstrates that even if the neurosurgical team uses PN, 4SYL, and SYNTAX (and potentially AN) stimuli that a patient processed accurately 1 day prior to surgery, there remains a risk of nonspecific trial failure in the operating room. The present findings suggest that when PN accuracy declines in the intraoperative context, the neurosurgical team should consider switching to a task with greater intraoperative accuracy, such as TR if it is anatomically relevant. Such anatomical regions reliably activated during text reading include the dominant inferior temporal gyrus, dominant inferior frontal gyrus, and dominant angular gyrus.[24] While the linguistic modalities tested in TR are inherently different from those in PN, the two tasks overlap in terms of visual processing and phonemic production, accounting for their relatively high anatomical concordance with the exception of the anterior temporal lobe, which is preferentially recruited in semantic processing during picture naming.[25] Therefore, awake mapping using TR may be more useful than an asleep image-guided resection in compatible regions if TR is the only task with reliable accuracy for an individual patient.[26,27]

The inclusion of the 4SYL task, in which the participant repeats a 4-syllable word presented via audio, may allow for awake language mapping in patients with poor vision and/or literacy (which would preclude PN and TR). Similarly, PN's counterpart in the auditory realm is AN. Here, for a given trial, a subject is expected to respond "banana" to the auditory stimulus "a curved fruit with a yellow peel." AN is thought to recruit more diffuse functional and anatomical networks compared to 4SYL, because it requires a component of decision-making.[28] This phenomenon is likely reflected by AN's lower task accuracy in both testing settings and its wider distribution of performance among study participants. The favorable sensitivity of AN and PN (and ability to capture more "natural" speech functions relative to simple text reading and repetition of audio) can be exploited in patients who perform at or near preoperative levels during prestimulation testing in the operating room. Further, because intraoperative performance of AN and PN is predictive of performance on the other tasks, patients with intact prestimulation accuracy on either AN or PN can likely proceed with stimulation mapping without the need for additional testing.

The present study also capitalized on a salient opportunity to determine whether patients who experience significant intraoperative declines in language task accuracy can be identified during preoperative testing. Unfortunately, with the exception of a single task (SYNTAX), preoperative task accuracy did not predict intraoperative task accuracy. This finding highlights the need to perform language testing following anesthetic withdrawal, but prior to stimulation mapping, to ensure that accuracy remains high. Indeed, without such a "retest," patients whose prestimulation task accuracy declines from their preoperative accuracy might undergo a more conservative extent of resection (due to increased false positives) during stimulation mapping.

Finally, the relatively modest percent changes from baseline in language task accuracy when averaged across the entire cohort highlight the important point that intraoperative language decline does not occur universally among patients presenting for awake craniotomy. Indeed, the majority of patients in this cohort had intact language function after 1) the minimum 20-minute postemergence waiting period during which anesthetics were completely withdrawn and 2) completing a series of objective "wakefulness tasks" according to established protocol. This provides the neurosurgical community with reassurance that, at least from a population level, intraoperative language assessments are valid behavioral measures when the aforementioned clinical practices are implemented prior to language testing. However, despite this modest population effect, it is clear that a subset of our cohort witnessed significant, unexpected declines in language function in some but not all tasks. Representative cases from this study are patients 4, 9, and 37, who witnessed at least 25% declines in syntax formation without accompanying deficits in text reading (Fig. 3). These patients, among others, drive the significant main effects demonstrated in this study and may benefit from undergoing stimulation mapping with tasks that can be performed at or near their baselines.

Study Limitations

The present study has two main limitations. First, we were unable to blind the task scorers to the testing settings. Additionally, while the scorers were instructed not to access records of prior language task performance for any patient, because we utilized the same scorers across testing sessions to maintain consistency in scoring strategy, scorers were not perfectly blinded to prior performance because they may have been directly involved in scoring those sessions. Second, 6 patients did not complete the full set of trial stimuli within each language task. However, we do not expect this to significantly affect our results given the relatively small number of missing data (608 trials at maximum) compared to the 19,448 total trials conducted in this study when pooled across pre- and intraoperative testing sessions (Supplemental Table 1).

Conclusions

The present findings indicate that accuracy on several language tasks is lower during intraoperative, prestimulation language testing than during preoperative testing 24 hours before surgery. This outcome suggests that some errors during language mapping may reflect reduced arousal or attention that is caused by intraoperative medications, rather than disrupted functioning of eloquent language cortex. However, we also observed relatively stable performance across testing environments for text reading. Thus, while visual and auditory object naming may provide greater sensitivity for anatomically localizing eloquent regions underlying natural speech, text reading should be considered as a viable alternative when prestimulation accuracy for the former tasks declines from high preoperative levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ABBREVIATIONS

AN	auditory object naming
PN	picture naming
QAB	quick aphasia battery
SYNTAX	production of syntactically intact sentences
TR	text reading
4SYL	repetition of 4-syllable words

References

1. Gogos AJ, Young JS, Morshed RA, et al. Awake glioma surgery: technical evolution and nuances. J Neurooncol. 2020; 147(3): 515–524. [PubMed: 32270374]

2. Hervey-Jumper SL, Li J, Lau D, et al. Awake craniotomy to maximize glioma resection: methods and technical nuances over a 27-year period. J Neurosurg. 2015; 123(2): 325–339.

3. Sanai N, Mirzadeh Z, Berger MS. Functional outcome after language mapping for glioma resection. N Engl J Med. 2008; 358(1): 18–27.

4. Szelényi A, Bello L, Duffau H, et al. Intraoperative electrical stimulation in awake craniotomy: methodological aspects of current practice. Neurosurg Focus. 2010; 28(2): E7.

5. Brown TJ, Brennan MC, Li M, et al. Association of the extent of resection with survival in glioblastoma: a systematic review and meta-analysis. JAMA Oncol. 2016; 2(11): 1460–1469.

6. Brown T, Shah AH, Bregy A, et al. Awake craniotomy for brain tumor resection: the rule rather than the exception? J Neurosurg Anesthesiol. 2013; 25(3): 240–247.

7. Groshev A, Padalia D, Patel S, et al. Clinical outcomes from maximum-safe resection of primary and metastatic brain tumors using awake craniotomy. Clin Neurol Neurosurg. 2017; 157: 25–30.

8. Kim SS, McCutcheon IE, Suki D, et al. Awake craniotomy for brain tumors near eloquent cortex: correlation of intraoperative cortical mapping with neurological outcomes in 309 consecutive patients. Neurosurgery. 2009; 64(5): 836–846. [PubMed: 19404147]

9. 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.

10. Ushio Y, Kochi M, Hamada J, et al. Effect of surgical removal on survival and quality of life in patients with supratentorial glioblastoma. Neurol Med Chir (Tokyo). 2005; 45(9): 454–461. [PubMed: 16195644]

11. De Witte E, Satoer D, Colle H, et al. Subcortical language and non-language mapping in awake brain surgery: the use of multimodal tests. Acta Neurochir (Wien). 2015; 157(4): 577–588. [PubMed: 25585834]

12. Ibrahim GM, Bernstein M. Awake craniotomy for supratentorial gliomas: why, when and how? CNS Oncol. 2012; 1(1): 71–83. [PubMed: 25054301]

13. Ille S, Engel L, Albers L, et al. Functional reorganization of cortical language function in glioma patients-a preliminary study. Front Oncol. 2019; 9: 446. [PubMed: 31231608]

14. Rofes A, Spena G, Miozzo A, et al. Advantages and disadvantages of intraoperative language tasks in awake surgery: a three-task approach for prefrontal tumors. J Neurosurg Sci. 2015; 59(4): 337–349. [PubMed: 26159550]

15. Fernández Coello A, Moritz-Gasser S, Martino J, et al. Selection of intraoperative tasks for awake mapping based on relationships between tumor location and functional networks. J Neurosurg. 2013; 119(6): 1380–1394. [PubMed: 24053503]

16. Aabedi AA, Ahn E, Kakaizada S, et al. Assessment of wakefulness during awake craniotomy to predict intraoperative language performance. J Neurosurg. 2020; 132(6): 1930–1937.

17. Lee MH, Kim SH, Seoul HJ, et al. Impact of maximal safe resection on the clinical outcome of adults with craniopharyngiomas. J Clin Neurosci. 2012; 19(7): 1005–1008. [PubMed: 22595354]

18. Costello TG, Cormack JR. Anaesthesia for awake craniotomy: a modern approach. J Clin Neurosci. 2004; 11(1): 16–19. [PubMed: 14642359]

19. Wilson SM, Eriksson DK, Schneck SM, Lucanie JM. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. PLoS One. 2018; 13(2): e0192773. [PubMed: 29425241]

20. Duffau H Stimulation mapping of white matter tracts to study brain functional connectivity. Nat Rev Neurol. 2015; 11(5): 255–265. [PubMed: 25848923]

21. Middlebrooks EH, Yagmurlu K, Szaflarski JP, et al. A contemporary framework of language processing in the human brain in the context of preoperative and intraoperative language mapping. Neuroradiology. 2017; 59(1): 69–87. [PubMed: 28005160]

22. Duffau H, Moritz-Gasser S, Mandonnet E. A re-examination of neural basis of language processing: proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming. Brain Lang. 2014; 131: 1–10. [PubMed: 23866901]

23. Nossek E, Matot I, Shahar T, et al. Failed awake craniotomy: a retrospective analysis in 424 patients undergoing craniotomy for brain tumor. J Neurosurg. 2013; 118(2): 243–249. [PubMed: 23121432]

24. Martin A, Schurz M, Kronbichler M, Richlan F. Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. Hum Brain Mapp. 2015; 36(5): 1963–1981. [PubMed: 25628041]

25. Lau JKL, Humphreys GW, Douis H, et al. The relation of object naming and other visual speech production tasks: a large scale voxel-based morphometric study. Neuroimage Clin. 2015; 7: 463–475. [PubMed: 25685713]

26. Ghinda CD, Duffau H. Network plasticity and intraoperative mapping for personalized multimodal management of diffuse low-grade gliomas. Front Surg. 2017; 4: 3. [PubMed: 28197403]

27. Rofes A, Mandonnet E, de Aguiar V, et al. Language processing from the perspective of electrical stimulation mapping. Cogn Neuropsychol. 2019; 36(3–4): 117–139. [PubMed: 29996708]

28. Cervenka MC, Corines J, Boatman-Reich DF, et al. Electro-corticographic functional mapping identifies human cortex critical for auditory and visual naming. Neuroimage. 2013; 69: 267–276. [PubMed: 23274183]
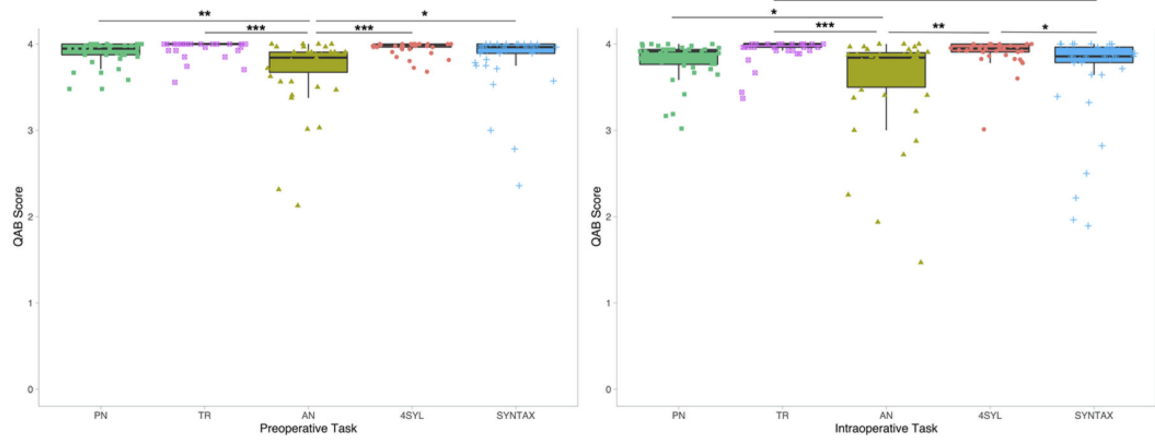
FIG. 1.

Comparisons in task accuracy in preoperative (**left**) and intraoperative (**right**) settings (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, adjusted for multiple comparisons) with median scores and interquartile ranges. Preoperatively, accuracy on AN was lower than accuracy on all other tasks. No other comparisons were statistically significant. TR and AN have lower variances among our subjects, with performance clustered at the top. AN has the largest variance. This signature holds true in the intraoperative setting. AN remains the most difficult task, although accuracy on SYNTAX decreased such that the difference between AN and SYNTAX was no longer statistically significant.
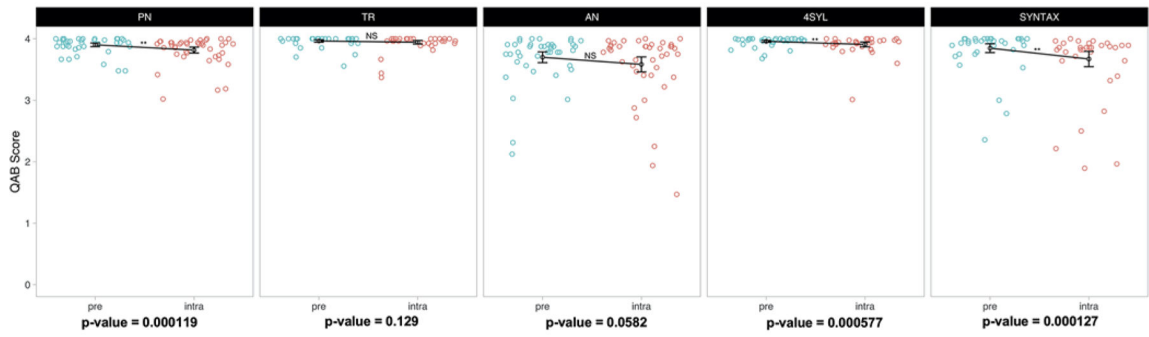
FIG. 2.

Paired nonparametric comparisons in preoperative and intraoperative performance grouped according to task. PN, 4SYL, and SYNTAX performance declined significantly in the intraoperative setting. The decline in PN replicates previous findings of a task-specific effect on intraoperative performance (Aabedi et al.[16]). The effect sizes of the PN and 4SYL declines are relatively modest compared with those of SYNTAX. There was a tendency toward decreased performance on AN in the intraoperative setting, although this did not reach statistical significance.
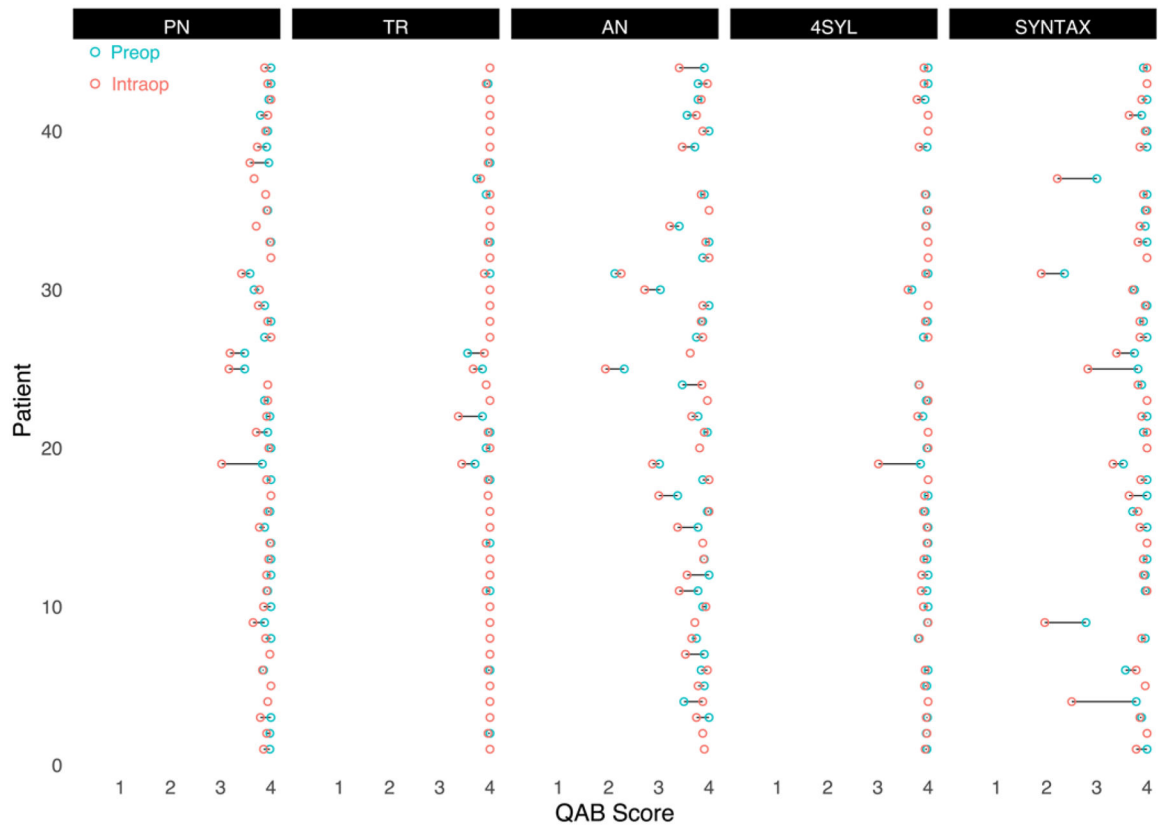
FIG. 3.
Dumbbell plot demonstrating change in performance of each patient by task. Most patients had only minor fluctuations in performance on TR and 4SYL. Declines in accuracy are more broadly distributed among the patient population in AN and SYNTAX.
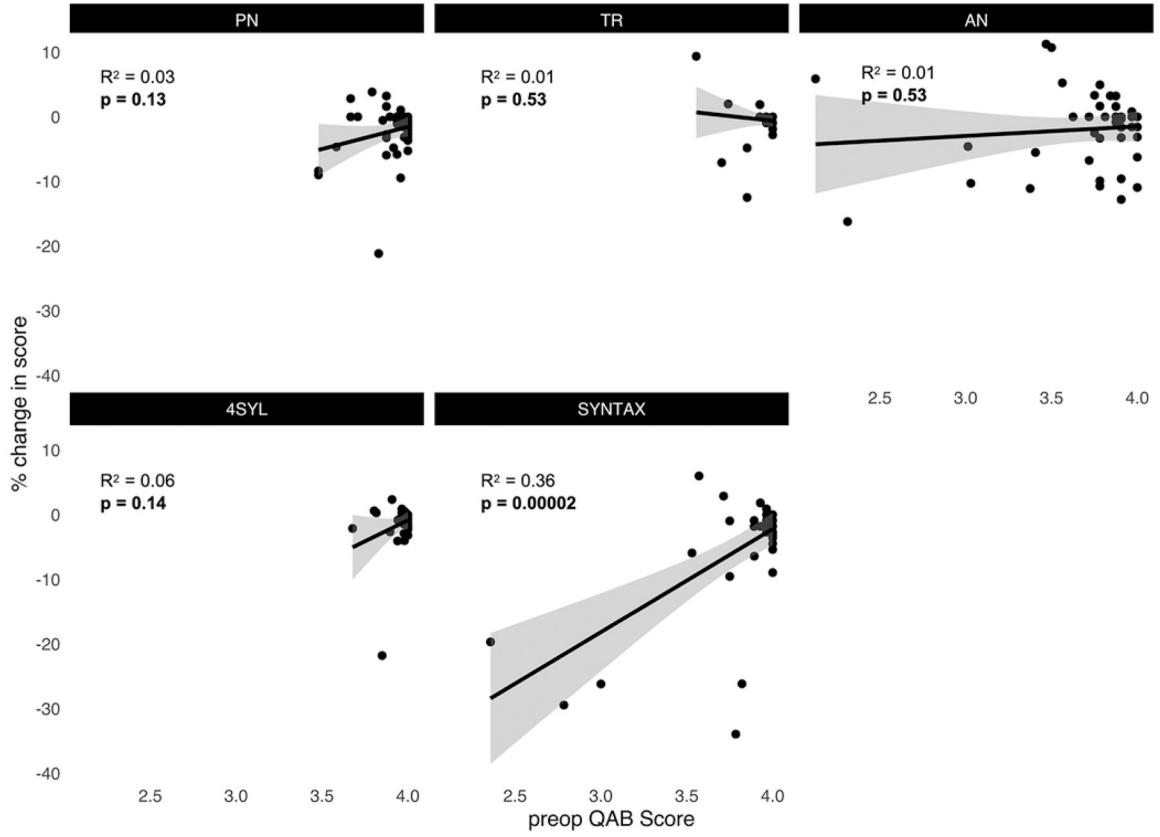
FIG. 4.

Linear regressions between preoperative (baseline) task accuracy and percent change in accuracy from baseline. There were no significant associations between baseline accuracy and subsequent changes in accuracy during intraoperative testing for PN, TR, AN, and 4SYL. High baseline accuracy on SYNTAX predicted stable performance in the operating room.

TABLE 1.

Clinical demographics

Characteristic	Value
Patients	44
Sex	
Female	15
Male	29
Mean age, yrs (range)	50.5 (19–81)
Handedness	
Left	2
Right	39
Unknown	3
Highest level of education	
Graduate	4
Undergraduate	7
Some college	2
High school	3
Unknown	28
Employment status	
Employed	15
Not employed	14
Unknown	15
Tumor laterality	
Left hemisphere	41
Right hemisphere	3
Predominant tumor location	
Frontal	17
Parietal	9
Temporal	16
Insular	2
Reoperation	6
Pathology	
WHO grade	
II	10
III	12
IV	19
Metastasis	2
Cavernous malformation	1

TABLE 2.

Univariate comparisons between intraoperative tasks

	PN	TR	AN	4SYL	SYNTAX
PN	—	0.0002	0.00003	0.000004	0.0004
TR	0.0002	—	0.005	0.000007	0.147
AN	0.00003	0.005	—	0.002	0.00001
4SYL	0.000004	0.000007	0.002	—	0.71
SYNTAX	0.0004	0.147	0.00001	0.71	—

All values are p values.