Int J CARS
DOI 10.1007/s11548-015-1160-y

ORIGINAL ARTICLE

Operator experience determines performance in a simulated computer-based brain tumor resection task

Terrell Holloway · Zachary S. Lorsch · Michael A. Chary · Stanislaw Sobotka · Maximillian M. Moore · Anthony B. Costa · Rolando F. Del Maestro · Joshua Bederson

Received: 5 September 2014 / Accepted: 10 February 2015

Abstract
Purpose Develop measures to differentiate between experienced and inexperienced neurosurgeons in a virtual reality brain surgery simulator environment.
Methods Medical students ($n = 71$) and neurosurgery residents ($n = 12$) completed four simulated Glioblastoma multiforme resections. Simulated surgeries took place over four days with intermittent spacing in between (average time between surgeries of 4.77 ± 0.73 days). The volume of tumor removed (cc), volume of healthy brain removed (cc), and instrument path length (mm) were recorded. Additionally, surgical effectiveness (% tumor removed divided by % healthy brain removed) and efficiency (% tumor removed divided by instrument movement in mm) were calculated. Performance was compared (1) between groups, and (2) for each participant over time to assess the learning curve. In addition, the effect of real-time instruction ("coaching") was assessed with a randomly selected group of medical students.
Results Neurosurgery residents removed less healthy brain, were more effective in removing tumor and sparing healthy brain tissue, required less instrument movement, and were more efficient in removing tumor tissue than medical students. Medical students approached the resident level of performance over serial sessions. Coached medical students showed more conservative surgical behavior, removing both less tumor and less healthy brain. In sum, neurosurgery residents removed more tumor, removed less healthy brain, and required less instrument movement than medical students. Coaching modified medical student performance.
Conclusions Virtual Reality brain surgery can differentiate operators based on both recent and long-term experience and may be useful in the acquisition and assessment of neurosurgical skills. Coaching alters the learning curve of naïve inexperienced individuals.

Keywords Simulator · Virtual reality · Training · NeuroTouch · Coaching · Glioblastoma multiforme tumor

Introduction

The consequences of errors during brain surgery can be severe [1]. Using surgical simulators, surgeons may practice in a virtual reality (VR) environment before performing operations in the physical world. This could theoretically reduce procedural errors, especially during training. Several brain surgery simulators exist [2–4]; however, there are few studies validating their use as an effective training tool [5].

A VR simulator that is integrated into a formal surgical training curriculum would advance the training and evalua-

Terrell Holloway and Zachary S Lorsch have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11548-015-1160-y) contains supplementary material, which is available to authorized users.

T. Holloway · Z. S. Lorsch · M. A. Chary · S. Sobotka · M. M. Moore · A. B. Costa · J. Bederson
Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

R. F. Del Maestro
Department of Neurology and Neurosurgery, Brain Tumor Research Centre, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

J. Bederson (✉)
Department of Neurosurgery, Mount Sinai Medical Center, 1468 Madison Avenue, Annenberg 8 – 28, New York, NY 10029, USA
e-mail: Joshua.Bederson@mssm.edu; joshua.bederson@mountsinai.org

tion of neurosurgical residents by helping residents develop surgical skills without risk for surgical errors in patients [6,7]. A simulator could also provide training at any time and provide experience for residents and students that extend beyond the available patient population. Furthermore, a validated VR simulator could be used to objectively assess specific components of surgical technique, potentially identifying strengths, weaknesses, and areas for improvement for experienced and inexperienced surgeons alike, ultimately resulting in improved patient care and surgical outcomes [6].

For VR technology to be formally incorporated into the training of neurosurgeons, its assessments must be both valid and reliable. More specifically, operating in the VR environment of the simulator must feel like operating on a patient in physical reality (face validity), the simulator must consistently record data on important aspects of surgical performance (content validity), the measurements made by the simulator must be accurate and the intended variables (construct validity), and data obtained from the simulator must correlate with data collected with other validated measures of surgical performance [8]. In terms of reliability, the user experience must be consistent each time and, without a change in surgical skill, performance on the simulator should not change.

Outside of neurosurgery, existing surgery simulators appear to have relatively realistic interfaces and target different surgical techniques that need to be learned by surgical residents [9–11]. As such, some components of face and content validity are inherent. However, as brain surgery simulators are relatively new, more detailed assessment of face validity, construct validity, concurrent validity, and reliability need to be performed [4]. In particular, while some studies have been able to differentiate skill based on output measures from a brain surgery simulator [12], few surgical scenarios and output measures that are directly applicable to clinical ability have been examined.

While most simulators provide basic output data encompassing measures such as force exerted by the operator, a quantification of instrument movement, and an assessment of how successful the VR operation was (such as how much tumor tissue was removed), recent work indicates that more detailed assessments, such as compound measures encompassing multiple variables, may be better suited to identify differences in surgical performance between groups [13]. In this study, Azarnoush et al. [13] found that attending neurosurgeons remove the same or more tumor and less healthy brain, complete surgeries more quickly, activate the foot pedal less, and move less, than neurosurgery residents during simulated resection of a brain tumor. Furthermore, the authors noticed significant differences between attending neurosurgeons and residents on more difficult simulations, where the attending physicians, but not the residents, showed more conservative surgical behavior. Similarly, this study found that while both groups spent most of the surgery applying low levels of force, residents applied high force levels more often than attending neurosurgeons. Since any valid brain surgery simulator would be expected to show differences between more experienced (attending physicians) and less experienced (residents) surgeons, data such as these are particularly useful in the validation of surgical simulators. However, this study is limited in that the data were obtained from a single novice and a single expert surgeon. While larger simulation studies exist for other fields [14–16], no such study of sufficient sample size exists for neurosurgery.

The aim of the current study is to determine whether performance on a VR simulation of a Glioblastoma multiforme (GBM) reflects surgical skill. This particular simulation was chosen since GBM is the most prevalent malignant brain tumor [17] and requires both substantial manual dexterity and clinical reasoning (most notably, balancing removal of tumor and healthy brain) for effective resection. While validation of a surgical simulators requires establishment of face validity, content validity, and construct validity [18,19], we chose to focus on construct validity since this approach is more objective to the alternatives (i.e., asking surgeons if the simulator feels realistic) in that it utilizes measures recorded by the simulator (such as force employed and hand movement) that would be difficult for even an experienced operator to self-assess. Furthermore, this study also expands previous work by measuring clinically relevant variables in longer surgical tasks, assessing learning curve data by repeating the same surgical task on different days to evaluate changes in performance over time, and evaluating the effect of coaching on improving surgical performance [12,13]. In non-neurosurgical simulations, coaching and repeated trials has been shown to improve outcomes and decrease operative error [20,21]. Identification of similar findings for brain surgery simulators would help to promote the use of these simulators as a training tool.

Within this study, we aim to identify specific measures that can be utilized to differentiate performance on a VR brain surgery GBM resection task and investigate how repeat performance, as well as coaching, can affect surgical skill in a VR environment. The results of this study can potentially help to validate a virtual GBM resection module as a useful brain surgery simulation and, more broadly, show that VR neurosurgical simulation can be used to effectively differentiate operators with different levels of real-world surgical experience. Finally, we aim to show that VR brain surgery simulation may be useful for improving surgical performance in inexperienced users, thus strengthening the case for the use of VR brain surgery simulation as a training tool.

Materials and methods/case material

NeuroTouch

The NeuroTouch simulator (www.neurotouch.ca) created by the National Research Council (NRC) of Canada was developed to simulate a wide array of neurosurgical procedures including ventriculostomy, endoscopic nasal navigation, and brain tumor resection [4,12,18,19,22]. Essential techniques such as hemostasis and microdissection are simulated in procedure-based simulations and included in training modules that the NeuroTouch provides.

The physical setup, VR interface, and simulated environment were constructed similar to previous studies [12,13] (Fig. 1, Supplementary Video 1). In brief, the NeuroTouch is a computer-based VR simulator environment that provides real-time haptic feedback via a physical user interface. Several different modules, including the GBM tumor utilized in this study, can be selected for simulation. The surgical operator stands in front of an adjustable-height computer screen, viewing the image on the screen through a stereoscope, which generates a 3D visualization (Fig. 1a). An auxiliary display is included for observers. At arms height an open plastic head and two haptic devices are fixed to a table (Fig. 1a). Various surgical tools can be attached to the haptic devices, which can be controlled via a foot pedal and adjusted with output knobs built into the table. The positions of the tools in space are assessed by the haptic devices and displayed on the screen in real time (Fig. 1b). An empty space (simulating a craniotomy) in the plastic head indicates the surgical field and corresponds to the virtual surgical field as it appears on the screen (Fig. 1c).

The NeuroTouch platform provides sensory feedback to the user in order to realistically simulate the surgical environment. The haptic devices provide force feedback (resistance) that differs based on the tool being utilized, type of tissue being explored, area of the tissue, and simulator settings (Fig. 1c). In addition to the appearance of the tools on the screen within the open surgical field, a dialog box appears when forces exceed a pre-defined threshold informing the user that the current force is excess. Audio cues are limited to unchanging heart and respirator sounds.

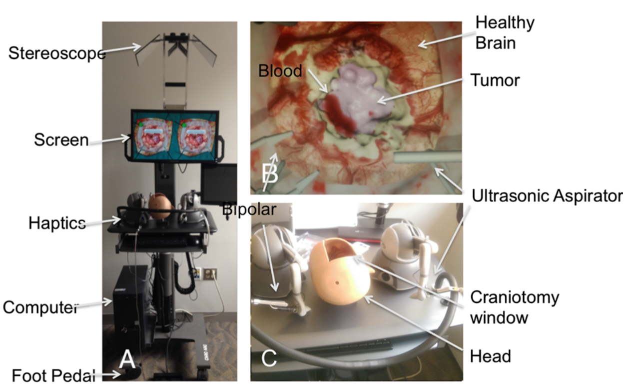

Fig. 1 NeuroTouch surgical simulator. **a** All hardware components of NeuroTouch surgical simulator. **b** Virtual representation of craniotomy. **c** Haptic hardware components including head as reference for surgical space

Simulator task

All simulations were completed on the NeuroTouch simulator. Participants were instructed to resect the standard GBM tumor provided by the NeuroTouch version 8/9/13. Our procedure resembles previous studies because it requires dexterity in both hands to remove tumor tissue while preserving surrounding healthy brain tissue [12,13]. We chose the GBM resection because we found it more intuitive and realistic than the alternative tumor module, the meningioma.

At the start of each session, each participant was given a one-page description of the task that explained how to identify tumor and healthy brain tissue, in addition to how to utilize the bipolar and aspirator tools necessary for the resection. After the participant confirmed that he/she understood the prompts, they were given a maximum of 10 min to resect the virtual tumor. Participants did not have to perform a skin incision, craniotomy, dural dissection, or incision into the brain to reach the tumor; as such procedures are beyond the capabilities of the current version of the NeuroTouch. Each participant performed the same GBM resection four times on four different days (minimum timing between sessions was 24 h; however, participant was allowed to set times based on their own personal schedule). No participants were allowed to practice on the simulator prior to or in between the four recorded sessions.

The specific surgical environment can be observed in Supplementary Video 1. The goal of the procedure was to remove as much of the virtual GBM as possible with minimal removal of the surrounding healthy brain. At the start of the simulation, participants were presented with an exposed GBM tumor surrounded by normal appearing brain tissue. In order to resect the tumor, participants were briefed on the task and NeuroTouch environment. The ultrasonic aspirator tool was attached to the haptic device corresponding to their non-dominant hand and the bipolar tool was attached to the haptic device corresponding to their dominant hand. In addition, participants were briefed on the control of those instruments using the on-table knob controls and foot pedal. As the participants resected the tumor, tissue was removed from the simulation on the screen in real time. In addition, removal of certain areas of the tumor or healthy brain resulted in bleeding and required cauterization.

Study population and recruitment

This study was conducted at The Icahn School of Medicine at Mount Sinai. Medical students ($n = 71$) and neurosurgi-

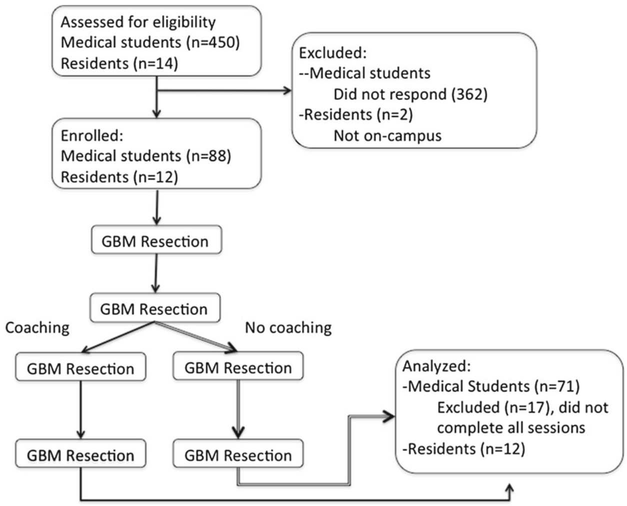

Fig. 2 CONSORT-style diagram. Enrollment of medical students and residents. Both medical students and residents were in the *non-coached* group (*double arrows*). Only medical students were assigned to the *coached* group (*single arrows*)

cal residents ($n = 12$) throughout all years of training were recruited to complete four simulated GBM resections on the NeuroTouch simulator (Fig. 2). Demographics of the study population are reflected in Table 1. The study was advertised throughout the school (flyers, emails, announcements). To encourage participation, medical students were compensated $100 upon completion of the study. Medical students were enrolled on a first come, first serve basis. All neurosurgical residents in the Mount Sinai Hospital Department of Neurosurgery participated in the study, except for two who were not on campus. Residents were not compensated. The institutional review board (IRB) of Mount Sinai Hospital determined that this study was exempt from IRB review.

Coaching

Coaching was performed during the surgery by a non-surgeon based on training and written instructions from an experienced attending neurosurgeon. This was done to simulate the intraoperative experience of trainees receiving feedback in the operating room. As such, since all residents had the actual experience of receiving feedback in the operating room from a neurosurgery attending in addition to previous experience operating on real patients, the residents were not coached. Every fifth medical student who enrolled received real-time instruction from a proctor during the final two GBM resections (coaching). The proctor, a medical student and co-investigator, TH, did not participate in the resection. He provided only verbal guidance. Coaching topics included not injuring the brain with excess force, minimizing the removal of healthy brain, and controlling bleeding through cauterization. Coaching was only provided on the last two sessions so that each subjects first two sessions provided an estimate of baseline performance.

Measures

Table 2 lists the variables recorded by the NeuroTouch simulator. All variables were recorded at a rate of 50 ± 5 Hz. These measures were examined for each subject in each session; however, variables that were non quantifiable (such as position), as well as variables that were less clinically significant, such as blood loss (because the location of bleeding was random) were not analyzed quantitatively and are not included.

Table 2 Variables

Simple	Composite	Derived values
Position (mm)	Volume (cc) of tumor removed—Vt	Tumor Removal Effectiveness $= \left(\frac{\mathrm{Vt}}{\mathrm{Vht}}\right)$
Velocity (mm/sec)	Volume (cc) of healthy tissue removed—Vht	Tumor Removal Efficiency $= \left(\frac{\mathrm{Vt}}{\mathrm{PL}}\right)$
Force (N)	Path length (mm) for each hand—PL	
Blood loss (cc)	Time spent (sec) exerting force >0.15 and 0.30 N	

A list of variables including the simple, complex, and derived variables analyzed in the study

Table 1 Demographics

	Total	Gender		Year		Handedness	
		Male	Female	Preclinical (M1&M2)	Clinical (M3&M4)	Left	Right
Medical students	71	41	30	46	25	4	67
	Total	Male	Female	Junior residents (PGY1-3)	Senior residents (PGY4-7)	Left	Right
Neurosurgical residents	12	8	4	6	6	1	11

A demographic table of the population cohort analyzed in this study

In addition to the variables output by the simulator, we derived two metrics (Table 2). To quantify the specificity of tumor removal, we calculated the ratio of the volume to tumor removed to the volume of healthy brain tissue removed. We term this ratio "tumor removal effectiveness." To quantify the specificity of movement, we calculated the volume of tumor removed per unit movement of the hand containing the aspirator. We termed this ratio "tumor removal efficiency." Given the effectiveness of compound measures (encompassing one or more variables) in previous studies of the NeuroTouch [12, 13], we expected these measures to better differentiate between resident and medical student performance than single output values alone.

Statistical analysis

The statistical approach in this paper is the analysis of short irregularly sampled time series and the determination of the correlation (1) between pairs of time series, and (2) between each time series and categorical or continuous demographic data. All graphs were created, and statistics and standard error of the mean (SEM) bars were calculated in Prism 6.0 by ZL and verified by MC using custom software written in Python. Each subject was labeled as either a medical student or a resident. Each medical student was further labeled as having been coached or not.

In order to ensure that there were no within-group differences, we first compared performance between junior and senior residents. We found no difference (data not shown) between these two groups so residents were treated as a single cohort ($n = 12$). Similarly, we examined the possibility of analyzing the data in terms of less experienced (medical students and junior residents) and more experienced (senior residents) operators, but since medical students and junior residents showed differences in performance (data not shown), we excluded this approach. Within the medical student group, coached and uncoached medical students were analyzed separately in the first two sessions to ensure that no confounding factors (i.e., video game proficiency, handedness, etc.) or baseline differences influenced the data after coaching began.

Results

Factors that distinguish medical students and neurosurgery residents in simulated brain surgery

Medical students removed significantly more tumor volume in the fourth session as compared with the first, 74.4 ± 12.8 % compared with 44.9 ± 22 %, respectively (Mann–Whitney test; $U = 1298$, $p < 0.001$; Fig. 3a). Neurosurgical residents removed the same amount of tumor in all sessions (Kruskal–Wallis test; $K = 4.921$, $p = 0.178$; Fig. 3a). For each session, there was no significant difference in tumor removed between residents and medical students.

Medical students removed more healthy brain in the fourth session as compared with the first, 8.5 ± 2.3 % compared with 5.9 ± 1.2 %, respectively (Friedman ANOVA; $Q = 19.94$, $p = 0.0002$; Fig. 3b). Neurosurgery residents removed a similar amount of brain over all four sessions 1.4–4.15 %, (Kruskal–Wallis test; $K = 2.563$, $p = 0.464$; Fig. 3b). Despite the trend for medical students to remove more healthy tissue in later sessions than the neurosurgical residents, there is no session in which there was a significant difference between the two groups (Mann–Whitney test; $U = 8.000$, $p = 1$).

We define tumor removal effectiveness as the ratio of volume of tumor removed to the volume of healthy brain tissue removed. This ratio quantifies how much healthy tissue a surgeon sacrifices to resect a given amount of tumor. Neurosurgery residents had higher tumor removal effectiveness than medical students for all four sessions (Mann–Whitney test; $U = 0$, $p = 0.0286$; Fig. 3c). Over four sessions, medical students improved in tumor removal effectiveness, from 6.28 ± 0.40 to 8.26 ± 0.35 (Friedman's one-way nonparametric ANOVA; $Q = 14.24$, $p = 0.0026$; Fig. 3c). Neurosurgery residents did not improve in tumor removal effectiveness (Friedman's one-way nonparametric ANOVA; $Q = 1.457$, $p = 0.6922$; Fig. 3c). For comparison, the tumor removal effectiveness for neurosurgical residents was 9.06 ± 1.399 in the first session and 9.91 ± 1.026 in the last session.

Medical students, whether coached or not, moved the aspirator more in the last session than the first one (Mann–Whitney test; $U = 1213$, $p < 0.0001$; Fig. 4a). Neurosurgery residents moved the aspirator a statistically indistinguishable amount in all sessions (Mann–Whitney test; $U = 25$, $p = 0.1632$; Fig. 4a). Finally, there was no statistical difference in the aspirator path length between medical students and residents.

Medical students spent a greater amount of time exerting excessive force in each subsequent session (Kruskal–Wallis test; $K = 21.16$, $p < 0.0001$; Fig. 4b). Neurosurgery residents did not (Kruskal–Wallis test; $K = 5.141$, $p = 0.1618$; Fig. 4b). Despite exerting increasing amounts of force, during no session did medical students exert significantly more force than residents (Mann–Whitney test; $U = 7$, $p = 0.8857$; Fig. 4b). We defined excessive force as that >0.3 N [12].

We defined tumor removal efficiency (as opposed to tumor removal *effectiveness*) as the amount of tissue removed divided by instrument movement to quantify how much a subject needed to move the instruments to remove tissue. Higher tumor removal efficiency scores indicate more economic movements. Medical students had similar tumor

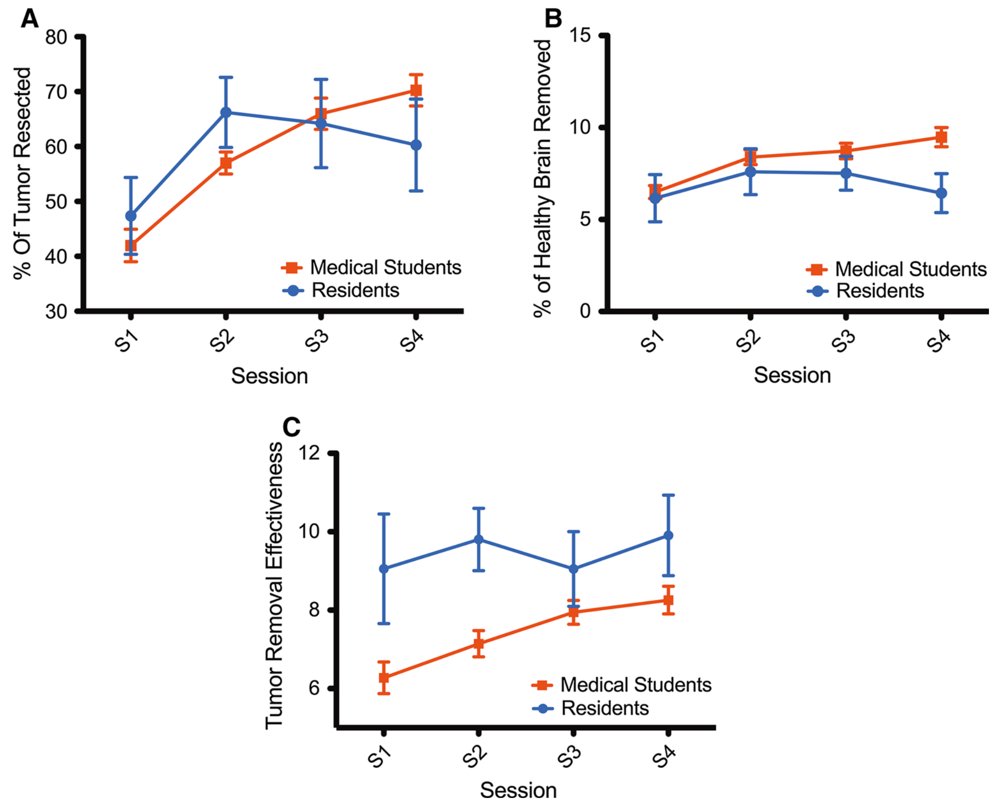

Fig. 3 Medical students and neurosurgery residents differ in their resection of tumor and removal of healthy brain. **a** Medical students removed significantly more tumor volume in the fourth session as compared with the first, $74.4 \pm 12.8\,\%$ compared with $44.9 \pm 22\,\%$, respectively (Mann–Whitney test; $U = 1298$, $p < 0.001$). Neurosurgical residents removed the same amount of tumor in all sessions (Kruskal–Wallis test; $K = 4.921$, $p = 0.178$). For each session, there was no significant difference in tumor removed between residents and medical students. **b** Medical students removed more healthy brain in the fourth session as compared with the first, $8.5 \pm 2.3\,\%$ compared with $5.9 \pm 1.2\,\%$, respectively (Friedman ANOVA; $Q = 19.94$, $p = 0.0002$). Neurosurgery residents removed a similar amount of brain over all four sessions 1.4–4.15 %, (Kruskal–Wallis test; $K = 2.563$, $p = 0.464$). Despite the trend for medical students to remove more healthy tissue in later sessions than the neurosurgical residents, there is no session in which there was a significant difference between the two groups (Mann–Whitney test; $U = 8.000$, $p = 1$). **c** Neurosurgery residents had higher tumor removal effectiveness than medical students for all four sessions (Mann–Whitney test; $U = 0$, $p = 0.0286$). Over four sessions, medical students improved in tumor removal effectiveness, from 6.28 ± 0.40 to 8.26 ± 0.35 (Friedman's one-way nonparametric ANOVA; $Q = 14.24$, $p = 0.0026$). Neurosurgery residents did not improve in tumor removal effectiveness (Friedman's one-way nonparametric ANOVA; $Q = 1.457$, $p = 0.6922$)

removal efficiency scores compared to residents (Mann–Whitney test; $U = 0$, $p = 0.286$; Fig. 4c), but improved over time (Kruskal–Wallis test; $K = 7.845$, $p = 0.0493$; Fig. 4c). For Neurosurgery residents, tumor removal efficiency did not improve with subsequent sessions (Kruskal–Wallis test; $K = 1.060$, $p = 0.787$, Fig. 4c).

The effect of coaching on medical student performance of a virtual brain surgery task

Coached medical students resected less tumor than uncoached ones (Mann–Whitney test; $U = 0.4$, $p < 0.001$; Fig. 5a), removing only $59.2 \pm 5.9\,\%$ of the tumor in the final session compared with $72.3 \pm 3.14\,\%$ for uncoached medical students. Coached medical students removed less healthy brain than uncoached ones (Mann–Whitney test; $U = 0.02$, $p < 0.001$; Fig. 5b), removing only $6.7 \pm 0.92\,\%$ compared with $10 \pm 0.57\,\%$ for uncoached students in the final session. However, this did not translate into a more effective surgery, as there were no differences between coached and uncoached medical students in tumor removal effectiveness (Kruskal–Wallis test; $K = 1.004$, $p = 0.216$; Fig. 5c). In addition, coaching reduced aspirator movement (Mann–Whitney test; $U = 0.02$, $p < 0.001$; upper panel, Fig. 6) and increased tumor removal efficiency for coached medical students (Mann–Whitney test; $U = 0$, $p < 0.0001$; lower panel, Fig. 6), though there was no significant difference in final session between coached and uncoached groups (Mann–Whitney test; $U = 0$, $p = 0.216$, lower panel, Fig. 6).

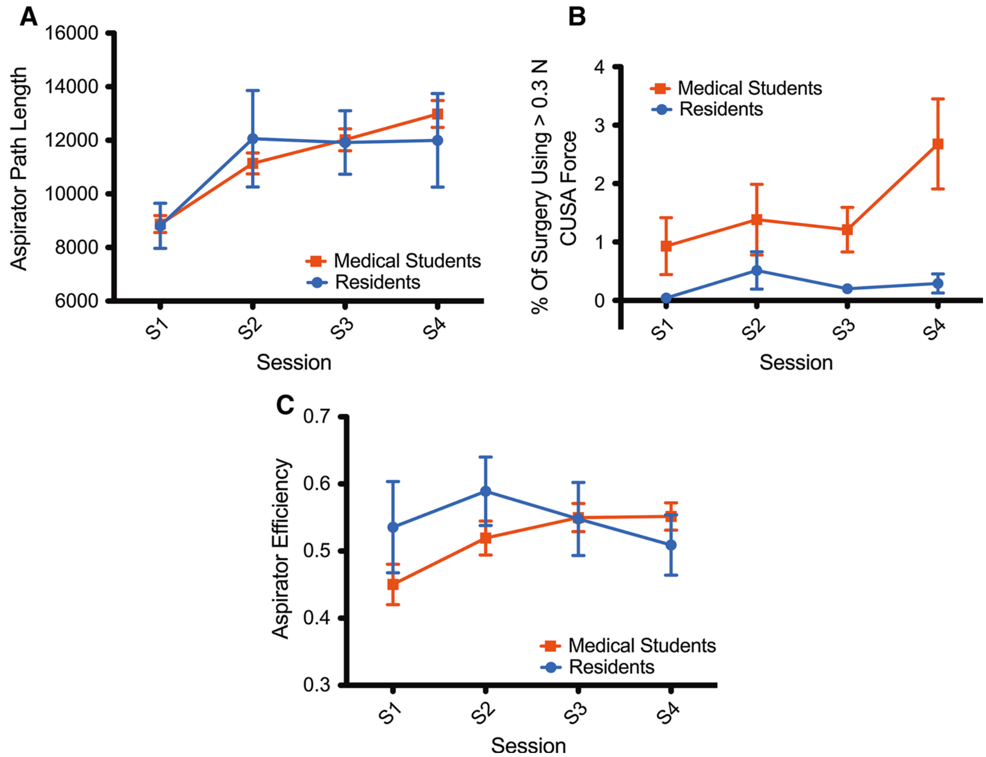

Fig. 4 Medical students and neurosurgery residents differ in the amount of movement and level of force utilized during surgery. **a** Medical students, whether coached or not, moved the aspirator more in the last session than the first one (Mann–Whitney test; $U = 1213$, $p < 0.0001$). Neurosurgery residents moved the aspirator a statistically indistinguishable amount in all sessions (Mann–Whitney test; $U = 25$, $p = 0.1632$; Fig. 4a). Finally, there was no statistical difference in the aspirator path length between medical students and residents. **b** Medical students spent a greater amount of time exerting excessive force in each subsequent session (Kruskal–Wallis test; $K = 21.16$, $p < 0.0001$). Neurosurgery residents did not (Kruskal–Wallis test; $K = 5.141$, $p = 0.1618$). Despite exerting increasing amounts of force, during no session did medical students exert significantly more force than residents (Mann–Whitney test; $U = 7$, $p = 0.8857$). **c** Medical students and residents had similar levels of efficiency (Mann–Whitney test; $U = 0$, $p = 0.286$) in the first and second session but improved over time (Kruskal–Wallis test; $K = 7.845$, $p = 0.0493$). For Neurosurgery residents, tumor removal efficiency did not improve with subsequent sessions (Kruskal–Wallis test; $K = 1.060$, $p = 0.787$)

Discussion

The use of virtual surgical simulators for neurosurgery is less firmly established than other subspecialties [2,23], particularly laparoscopic surgery [16,24–27]. This study provides evidence that performance on a virtual surgical simulator reflects neurosurgical proficiency in the physical world by demonstrating that medical students and neurosurgical residents perform differently on virtual brain surgery simulation. Our study also demonstrated that performance on a surgical simulator improves over time in medical students who are naïve and relatively inexperienced to brain surgery procedures, which suggests that virtual surgery simulation may be useful as an early training tool.

Previous studies demonstrated that medical students remove less tumor and are less efficient than neurosurgery residents in NeuroTouch simulations [4,12]. Gelinas et al. [12] found that medical students and residents removed different volumes of tumor, although they performed a slightly different task on the NeuroTouch, resecting a virtual meningioma as opposed to a virtual GBM, which was performed in our study. Our finding that students were both less effective and less efficient in removing tumor further supports these findings. Our finding that medical students but not neurosurgical residents improved in performance over four sessions extends prior findings by demonstrating that the difference between the surgical performance of residents and medical students persists over time but can be lessened by virtual surgery simulation.

Our data indicate that medical students' performance changes over time by removing more tissue, both tumor and healthy brain. However, tumor removal effectiveness also increased, which suggests that medical students can learn to remove less healthy brain per unit volume tumor. This finding suggests that medical students are initially focused on maximizing tumor resection and differentiates them from neurosurgical residents who know to balance the trade-off between collateral damage and tumor resection. A similar pattern

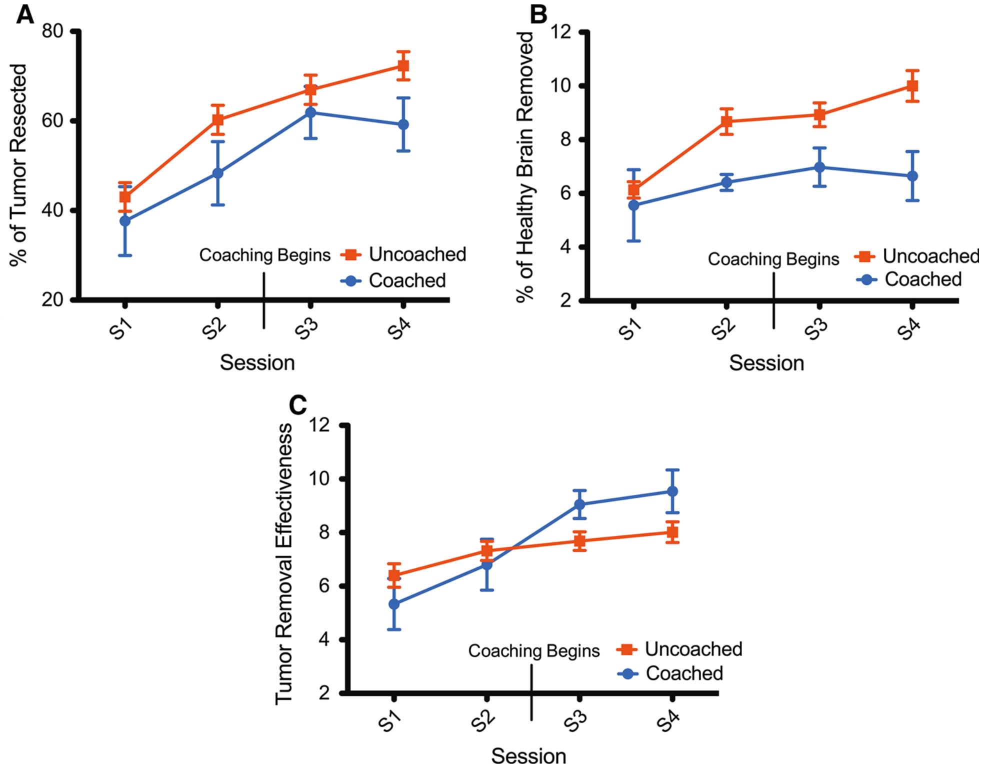

Fig. 5 Coached and uncoached medical students show different patterns of tumor resection and removal of healthy brain. **a** Coached medical students resected less tumor than uncoached ones (Mann–Whitney test; $U = 0.4$, $p < 0.001$). **b** Coached medical students removed less healthy brain than uncoached ones (Mann–Whitney test; $U = 0.02$, $p < 0.001$). **c** Despite guidance, coached and uncoached medical students showed no difference in their tumor removal effectiveness (Kruskal–Wallis test; $K = 1.004$, $p = 0.216$)

of maximizing tumor resection at the expense of healthy brain removed was found between residents and experts in Azarnoush et al. [13], where, although the volume of tumor removed was relatively similar, the expert surgeon usually removed less healthy brain [13]. As the tumor removal effectiveness for medical students increases over subsequent sessions, medical students acquire this skill with practice and begin to behave more like residents and experts. However, in order to remove more tumor, medical students also moved the aspirator more and exerted more force. Perhaps a study with longer sessions will show an inflection point where tumor removal efficiency and effectiveness move in tandem. This upward trend for tumor resection, removal of healthy brain, and tumor removal effectiveness did not occur for neurosurgery residents, who remove more from session one and do not change in performance over time.

In the current study, medical students but not residents improved, which raises concerns about the utility of this surgical simulation task for neurosurgery residents who have already acclimated a certain level of surgical experience. Perhaps the task was appropriately difficult for medical students but too easy for neurosurgical residents. It does stand to reason that trainees at different levels will develop different skills and further refinement of VR simulations with improved image segmentation, haptics, and graphics representations as well as incorporated complications such as bleeding, swelling, or hypotension will make simulators more realistic and may increase the difficulty for residents. However, it is also possible that residents did improve over time, just not on the measures that were being recorded. Unfortunately, the current data are limited in that we are unable to distinguish between these two possibilities.

Overall, we found that coaching medical students encouraged conservative surgical behavior. Coached students removed less tumor, but also less healthy brain. Coached medical students also showed less aspirator movement. This is consistent with other studies of non-neurosurgical simulators [15,28,29]. Future studies in which medical students are coached for more sessions may provide more insight into the effect of coaching on the acquisition of skills in a virtual surgery simulator.

Showing similar patterns of performance across many different neurosurgical procedures could strengthen these findings. Perhaps some procedures are more amenable to simulation than others, owing to the nature of the procedure or limitations of current surgical simulation hardware or soft-

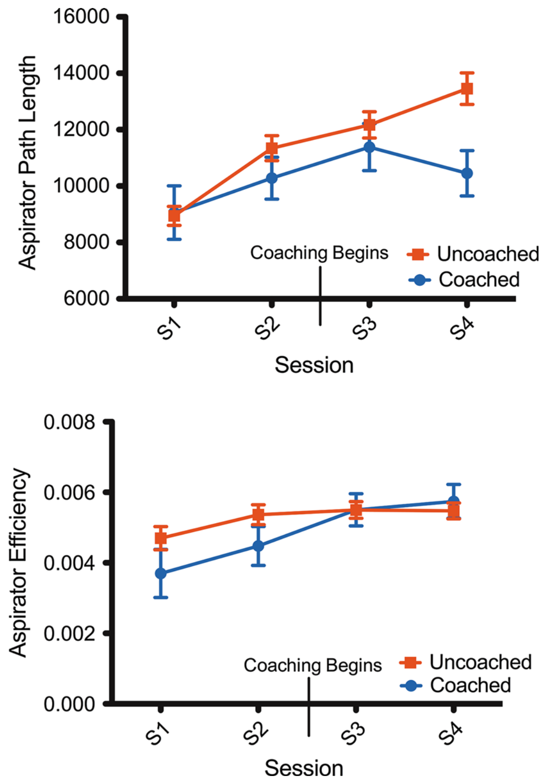

Fig. 6 Coached and uncoached medical students differ in their movement of the aspirator. *Upper* coaching reduced aspirator movement (Mann–Whitney test; $U = 0.02$, $p < 0.001$). *Lower* there was no difference in level of aspirator efficiency between coached and uncoached medical students. (Mann–Whitney test; $U = 0$, $p = 0.2157$). However, coaching increased aspirator efficiency between the first uncoached session and the last coached session (Mann–Whitney test; $U = 0$, $p = 0.0086$)

ware. In combination with further examination of measures that can distinguish neurosurgeons based on level of experience, and careful analysis of how repeated VR simulations affect surgical skill, establishing external validity of brain surgery simulators will combine with the current results to support the use of VR simulation in neurosurgery.

Conclusions

These data indicate that both long-term and recent experience influence proficiency in a simulated brain tumor resection task. Neurosurgery residents can be differentiated from medical students as they removed less healthy brain, were more effective in removing tumor and sparing healthy brain tissue, required less aspirator movement, and were more efficient in removing tumor tissue than medical students. Medical students learned rapidly, gradually approaching the resident level of performance, and coaching modified performance. These findings suggest that VR brain surgery may be useful in the acquisition and assessment of neurosurgical skills.

Acknowledgments We would like to thank the NRC for logistical support, Steven Philemond for coaching participants in the study, and Dr. Errol Gordon and Dr. E. J. Fernandez for their assistance with experimental design, the medical students and neurosurgery residents of the Icahn School of Medicine at Mount Sinai for their participation in our study, and all those who contributed to the study or manuscript preparation. Zachary Lorsch and Michael Chary are supported by the Medical Scientist Training Program of the Icahn School of Medicine at Mount Sinai (NIH Grant # T32 GM 007280).

Conflict of interest Terrell Holloway, Zachary Lorsch, Michael Chary, Stanislaw Sobotka, Maximilian Moore, Anthony Costa, and Rolando Del Maestro report no Conflicts of Interest. As stated in the disclosures, Dr. Joshua Bederson, is part of the Neurotouch Consortium advisory board without compensation.

Ethical standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study. The institutional review board (IRB) of Mount Sinai Hospital determined that this study was exempt from IRB review.

References

1. Aarabi B, Hesdorffer DC, Ahn ES, Aresco C, Scalea TM, Eisenberg HM (2006) Outcome following decompressive craniectomy for malignant swelling due to severe head injury. J Neurosurg 104:469–479. doi:10.3171/jns.2006.104.4.469
2. Robison RA, Liu CY, Apuzzo MLJ (2011) Man, mind, and machine: the past and future of virtual reality simulation in neurologic surgery. World Neurosurg 76:419–430. doi:10.1016/j.wneu.2011.07.008
3. Schmitt PJ, Agarwal N, Prestigiacomo CJ (2012) From planes to brains: parallels between military development of virtual reality environments and virtual neurological surgery. World Neurosurg 78:214–219. doi:10.1016/j.wneu.2012.06.014
4. Choudhury N, Gélinas-Phaneuf N, Delorme S, Del Maestro R (2012) Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. World Neurosurg 9–19. doi:10.1016/j.wneu.2012.08.022
5. Kockro RA (2013) Neurosurgery simulators-beyond the experiment. World Neurosurg 80:e101–e102. doi:10.1016/j.wneu.2013.02.017
6. Aoun SG, Bendok BR, Rahme RJ, Dacey RG, Batjer HH (2013) Standardizing the evaluation of scientific and academic performance in neurosurgery-critical review of the "h" Index and its variants. World Neurosurg 80:e85–e90. doi:10.1016/j.wneu.2012.01.052
7. Munz Y, Almoudaris AM, Moorthy K, Dosis A, Liddle AD, Darzi AW (2007) Curriculum-based solo virtual reality training for laparoscopic intracorporeal knot tying: objective assessment of the transfer of skill from virtual reality to reality. Am J Surg 193:774–783. doi:10.1016/j.amjsurg.2007.01.022
8. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ (2005) Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19:1523–1532. doi:10.1007/s00464-005-0384-2
9. Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G, Chandrasekhar R, Wilding GE, Guru KA (2010) Face validation of a novel robotic surgical simulator. Urology 76:357–360. doi:10.1016/j.urology.2009.11.069

10. Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A (2009) Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. Urology 73:1288–1292. doi:10.1016/j.urology.2008.12.044
11. Lee JY, Mucksavage P, Kerbl DC, Hunyh VB, Etafy M, McDougall EM (2012) Validation study of a virtual reality robotic simulator-role as an assessment tool? J Urol 187:998–1002. doi:10.1016/j.juro.2011.10.160
12. Gélinas-Phaneuf N, Choudhury N, Al-Habib AR, Cabral A, Nadeau E, Mora V, Pazos V, Debergue P, DiRaddo R, Del Maestro RF (2013) Assessing performance in brain tumor resection using a novel virtual reality simulator. Int J Comput Assist Radiol Surg. doi:10.1007/s11548-013-0905-8
13. Azarnoush H, Alzhrani G, Winkler-Schwartz A, Alotaibi F, Gelinas-Phaneuf N, Pazos V, Choudhury N, Fares J, DiRaddo R, Del Maestro RF (2014) Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. Int J Comput Assist Radiol Surg. doi:10.1007/s11548-014-1091-z
14. Paschold M, Schröder M, Kauff DW, Gorbauch T, Herzer M, Lang H, Kneist W (2011) Virtual reality laparoscopy: which potential trainee starts with a higher proficiency level? Int J Comput Assist Radiol Surg 6:653–662. doi:10.1007/s11548-010-0542-4
15. Selvander M, Åsman P (2012) Virtual reality cataract surgery training: learning curves and concurrent validity. Acta Ophthalmol 90:412–417. doi:10.1111/j.1755-3768.2010.02028.x
16. Iwata N, Fujiwara M, Kodera Y, Tanaka C, Ohashi N, Nakayama G, Koike M, Nakao A (2011) Construct validity of the LapVR virtual-reality surgical simulator. Surg Endosc 25:423–428. doi:10.1007/s00464-010-1184-x
17. Ohgaki H, Kleihues P (2005) Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. J Neuropathol Exp Neurol 64:479–489
18. Gélinas-Phaneuf N, Del Maestro RF (2013) Surgical expertise in neurosurgery: integrating theory into practice. Neurosurgery 73(Suppl 1):30–38. doi:10.1227/NEU.0000000000000115
19. Delorme S, Laroche D, DiRaddo R, Del Maestro RF (2012) NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. Neurosurgery 71:32–42. doi:10.1227/NEU.0b013e318249c744
20. Cole SJ, Mackenzie H, Ha J, Hanna GB, Miskovic D (2014) Randomized controlled trial on the effect of coaching in simulated laparoscopic training. Surg Endosc 28:979–986. doi:10.1007/s00464-013-3265-0
21. Teitelbaum EN, Soper NJ, Santos BF, Rooney DM, Patel P, Nagle AP, Hungness ES (2014) A simulator-based resident curriculum for laparoscopic common bile duct exploration. Surgery 156:880–887, 890–893. doi:10.1016/j.surg.2014.06.020
22. Rosseau G, Bailes J, del Maestro R, Choudhury N, Comas O, Debergue P, De Luca G, Hovdebo J, Jiang D, Laroche D, Neubauer A, Pazos V, Thibault F, DiRaddo R (2013) The development of a virtual simulator for training neurosurgeons to perform and perfect endoscopic endonasal transsphenoidal surgery. Neurosurgery 73(Suppl 1):85–93. doi:10.1227/NEU.0000000000000112
23. Neubauer A, Wolfsberger S (2013) Virtual endoscopy in neurosurgery: a review. Neurosurgery 72(Suppl 1):97–106. doi:10.1227/NEU.0b013e31827393c9
24. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, Ramel S, Smith CD, Arvidsson D (2007) Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 193:797–804. doi:10.1016/j.amjsurg.2006.06.050
25. Badurdeen S, Abdul-Samad O, Story G, Wilson C, Down S, Harris A (2010) Nintendo Wii video-gaming ability predicts laparoscopic skill. Surg Endosc 24:1824–1828. doi:10.1007/s00464-009-0862-z
26. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J (2003) Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc 17:1082–1085. doi:10.1007/s00464-002-9176-0
27. Kolozsvari NO, Andalib A, Kaneva P, Cao J, Vassiliou MC, Fried GM, Feldman LS (2011) Sex is not everything: the role of gender in early performance of a fundamental laparoscopic skill. Surg Endosc 25:1037–1042. doi:10.1007/s00464-010-1311-8
28. Schlickum MK, Hedman L, Enochsson L, Kjellin A, Fellander-Tsai L (2009) Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. World J Surg 33:2360–2367. doi:10.1007/s00268-009-0151-y
29. Boyle E, Kennedy A-M, Traynor O, Hill ADK (2011) Training surgical skills using nonsurgical tasks-can Nintendo Wii™ improve surgical performance? J Surg Educ 68:148–154. doi:10.1016/j.jsurg.2010.11.005