A systematic review of virtual reality for the assessment of technical skills in neurosurgery

*Justin Chan, BS; 1 Dhiraj J. Pangal, BS; 1 Tyler Cardinal, BS; 1 Guillaume Kugener, MEng; 1 Yichao Zhu, MS; 1 Arman Roshnai, 1 Nicholas Markarian, BS; 1 Aditya Sinha, BS; 1 Anima Anandkumar, PhD; 2 Andrew Hung, MD; 3 Gabriel Zada, MD, MS; 1 and Daniel A. Donoho, MD 4

1 USC Department of Neurosurgery, Keck School of Medicine of the University of Southern California, Los Angeles, California; 2 Computing + Mathematical Sciences, California Institute of Technology, Pasadena, California; 3 USC Department of Urology, Keck School of Medicine of the University of Southern California, Los Angeles, California; and 4 Texas Children’s Hospital, Baylor College of Medicine, Houston, Texas

OBJECTIVE Virtual reality (VR) and augmented reality (AR) systems are increasingly available to neurosurgeons. These systems may provide opportunities for technical rehearsal and assessments of surgeon performance. The assessment of neurosurgeon skill in VR and AR environments and the validity of VR and AR feedback has not been systematically reviewed.

METHODS A systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was conducted through MEDLINE and PubMed. Studies published in English between January 1990 and February 2021 describing the use of VR or AR to quantify surgical technical performance of neurosurgeons without the use of human raters were included. The types and categories of automated performance metrics (APMs) from each of these studies were recorded.

RESULTS Thirty-three VR studies were included in the review; no AR studies met inclusion criteria. VR APMs were categorized as either distance to target, force, kinematics, time, blood loss, or volume of resection. Distance and time were the most well-studied APM domains, although all domains were effective at differentiating surgeon experience levels. Distance was successfully used to track improvements with practice. Examining volume of resection demonstrated that attending surgeons removed less simulated tumor but preserved more normal tissue than trainees. More recently, APMs have been used in machine learning algorithms to predict level of training with a high degree of accuracy. Key limitations to enhanced-reality systems include limited AR usage for automated surgical assessment and lack of external and longitudinal validation of VR systems.

CONCLUSIONS VR has been used to assess surgeon performance across a wide spectrum of domains. The VR environment can be used to quantify surgeon performance, assess surgeon proficiency, and track training progression. AR systems have not yet been used to provide metrics for surgeon performance assessment despite potential for intraoperative integration. VR-based APMs may be especially useful for metrics that are difficult to assess intraoperatively, including blood loss and extent of resection.

https://thejns.org/doi/abs/10.3171/2021.5.FOCUS21210

KEYWORDS virtual reality; augmented reality; technical assessment

VIRTUAL reality (VR) and augmented reality (AR) systems have been proposed as methods of surgical performance improvement for decades, and have grown in complexity with recent iterations.1,6 VR environments are fully computer generated without direct connection to the physical space, whereas AR environments contain input from the physical space that is augmented by adding a sensory layer (such as graphics). The earliest VR systems were developed in the late 2000s, including a ventriculostomy environment using ImmersiveTouch and a brain tumor resection environment using NeuroTouch (subsequently named NeuroVR).2–4 Neurosurgery was an

ABBREVIATIONS AI = artificial intelligence; APM = automated performance metric; AR = augmented reality; EVD = external ventricular drain; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SVM = support vector machine; VR = virtual reality.

SUBMITTED March 31, 2021. ACCEPTED May 19, 2021.

INCLUDE WHEN CITING DOI: 10.3171/2021.5.FOCUS21210.

* J.C. and D.J.P. contributed equally to this work.
early adopter of AR technology due to the importance of stereotaxy for neuronavigation, with the earliest AR systems developed in the late 1990s.3–7

Recent VR systems have evolved to include a wide array of pathologies and surgical scenarios, allowing them to be used for trainee practice in low-stakes environments.1 In addition, modern VR systems are often integrated simulators that incorporate complex audiovisual cues, haptic interfaces, and benchtop systems, a clear improvement in sophistication from earlier VR headset paradigms.2–4 AR systems have evolved to provide real-time intraoperative visual navigation and allow for improved perioperative planning.3 Both systems have been effective methods of surgical rehearsal. These systems additionally have built-in sensors and monitors that can provide real-time or aggregate information on instrument movement, tool tracking, and other automated performance metrics (APMs), which are either not available within traditional operative neurosurgical settings or are difficult to quantify (e.g., blood loss, volume of tumor removed).4,5

Although AR and VR systems have grown in popularity and complexity, their ability to generate APMs that accurately assess surgeons’ technical skill remains unknown. The purpose of this systematic review was to assess the ability of these systems to quantitatively determine surgeon performance, compare surgeon performances, or demonstrate progression toward surgical competency. We additionally discuss pitfalls common to VR and AR systems and review future directions for the incorporation of these tools within neurosurgical training programs.

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines were used. We searched MEDLINE and PubMed for articles published between January 1, 1990, and February 2, 2021, using MeSH terms organized around the following key concepts: “augmented reality,” “virtual reality,” “surgical assessment,” and “neurosurgery.”

The full search strategy for the MEDLINE/PubMed query can be found below.

**Search Terms**

(((computer vision[MeSH Terms]) OR (artificial intelligence[MeSH Terms]) OR (machine learning[MeSH Terms]) OR (deep learning[MeSH Terms]) OR (automated pattern recognition[MeSH Terms]) OR (time and motion studies[MeSH Terms]) OR (video assisted surgery[MeSH Terms]) OR (eye-tracking technology[MeSH Terms]) OR (kinematics[MeSH Terms]) OR (virtual reality[MeSH Terms]) OR (augmented reality[MeSH Terms]) OR (computer vision)) AND (artificial intelligence) OR (machine learning)) OR (deep learning) OR (automated pattern recognition) OR (time and motion studies) OR (video assisted surgery) OR (eye-tracking technology) OR (kinematics) OR (virtual reality) OR (augmented reality) OR (tracking) OR (video data) OR (force detection) AND ((neurosurgery[MeSH Terms]) OR (neurosurgical procedure[MeSH Terms]) OR (neurosurgery))

**Results**

After removal of duplicates and studies published before 1990, 2107 studies remained. After abstract and title screening, 48 articles from reviewer 1 and 47 articles from reviewer 2 remained for full-text analysis. Following full-text analysis and application of inclusion/exclusion criteria, 33 VR studies were included for review. There were no studies exclusively using AR that met inclusion criteria. Of the 33 articles, 2 were published between 2005 and 2009, 7 between 2010 and 2014, 19 between 2015 and 2019, and 5 between 2020 and the time of this writing.

The most common VR system used was NeuroVR (CAE, Inc.), formerly named NeuroTouch, used in 15 studies (Fig. 2). This system serves as a VR benchtop simulator with head and haptic feedback—enabled surgical instruments for the following pathologies: brain tumor removal, hemilaminectomy, and aneurysm repair.8 The ImmersiveTouch system (ventriculostomy simulation) was...
FIG. 1. PRISMA flow diagram with number of articles remaining during the systematic review process.

FIG. 2. Various NeuroTouch/NeuroVR setups. A–D: Figure 2A–D, respectively, were reprinted from Sawaya R, Alsideiri G, Bugdadi A, et al. Development of a performance model for virtual reality tumor resections. J Neurosurg. 2018;131(1):192–200. Published with permission. E: Figure 2C was reprinted from Bajunaid K, Mullah MAS, Winkler-Schwartz A, et al. Impact of acute stress on psychomotor bimanual performance during a simulated tumor resection task. J Neurosurg. 2017;126(1):71–80. Published with permission.
used in 8 studies (Fig. 3). Other systems were used for 10 studies. Descriptions of all VR systems are found in Table 1. All of these models collect data on at least one of the major APM categories described in the Methods section. The number of studies included in each of the major APM categories is found in Table 2.

**Distance to Target**

Of 33 studies, 29 used the distance between an implanted device’s final placement and the surgeon’s intended target as a performance metric (Supplemental Table 1). Distance-to-target measurements have been used to assess performance during VR pedicle screw and ventricular catheter placements as well as mixed AR/VR pedicle cannulation and deep-seated tumor resection. In VR ventriculostomy procedures, distance between the foramen of Monro and catheter tip was not associated with location of the burr hole chosen. In contrast, increasing postgraduate year was significantly associated with decreased distance from the ideal entry point and target point in a percutaneous rhizotomy task. Repetition improved performance irrespective of training level, but deskilling sometimes occurred if repetitions were spaced far apart. Specifically, senior residents outperformed staff neurosurgeons in the number of attempts needed to place an external ventricular drain (EVD), likely due to senior residents placing a higher number of EVDs in the previous year despite significantly lower lifelong numbers. More detailed descriptions of these 29 studies can be found in Supplemental Table 1.

**Completion Time and Time-in-Phase**

Twenty-five of 33 studies measured time to complete a surgical task or phase of surgery across a variety of settings (Supplemental Table 2). Time plays an integral role not only in calculating kinematics but also in deriving complex metrics that use multiple categories of APMs to measure performance qualities (e.g., efficiency). One study found that all residents, regardless of postgraduate year, were able to conduct simulated ventriculostomies more quickly after practice and instruction. Another study found that time to complete a brain tumor resection task was longer in a resident participant than a faculty neurosurgeon. However, other studies were not able to differentiate surgical levels given time-based metrics alone.

**Surgical Tool Movement Metrics (kinematics)**

Kinematics are measures of instrument movement and position across time. Typical kinematics include velocity, acceleration, and jerk (the first derivative of acceleration). Seven studies used kinematics to measure neurosurgical performance (Supplemental Table 3); 6 examined kinetics...
TABLE 1. Distribution of the 33 studies by VR system

| VR System       | Description                                      | Procedure (no.)                                                                 | APM Domain                  | Distance | Time | Kinematics | Force | BL | Volume |
|-----------------|--------------------------------------------------|---------------------------------------------------------------------------------|----------------------------|----------|------|-----------|-------|-----|--------|
| NeuroVR*        | Benchtop simulator head; haptics-enabled instruments | Brain tumor resection (13); hemi-laminectomy; aneurysm repair                  |                            | ✓        | ✓    |           |       |     | ✓      |
| ImmersiveTouch  | Benchtop simulator head; haptics-enabled instruments | Ventriculostomy (5); spinal needle placement; pedicle screw placement; percutaneous rhizotomy |                            | ✓        | ✓    |       |       |     | ✓      |
| Sim-Ortho       | Video-based platform; haptic & auditory feedback; operators wear 3D glasses | Anterior cervical discectomy & fusion (2)                                      |                            | ✓        | ✓    | ✓        |       |     | ✓      |
| Own AR/VR†      | Abhari et al., 20159                              | Brain tumor resection                                                          |                            | ✓        | ✓    |       |       |     |         |
| Burström et al., 20191516 | Pedicle cannulation                             |                                                                            |                            | ✓        |       |       |       |     |         |
| Own VR          | Roitberg et al., 2013, 201517,18                  | Pedicle cannulation (2)                                                        |                            | ✓        | ✓    |       |       |     |         |
| Hooten et al., 201419 | Ventriculostomy                                 |                                                                            |                            | ✓        | ✓    |       |       |     |         |
| Cagiltay et al., 201714 | Brain tumor resection                          |                                                                            |                            | ✓        | ✓    |       |       |     |         |
| Heredia-Pérez et al., 201916 | Brain tumor resection                          |                                                                            |                            | ✓        | ✓    |       |       |     |         |
| Teodoro-Vite et al., 202012 | Aneurysm repair                                 |                                                                            |                            | ✓        | ✓    | ✓        |       |     |         |

BL = blood loss.
Descriptions for each VR system were outlined or referenced. The associated VR procedure with APMs examined was listed.
* Formerly known as NeuroTouch.
† For these studies, computer-generated information overlaid the real world (AR) with the option for a completely VR view or derived APMs from VR.

TABLE 2. APMs used to quantify surgeon performance from VR simulators, divided by APM domain

| APM Domain | No. of Studies | Pathologies (no.)                  | Total Participants* | Summarized Analysis |
|------------|----------------|------------------------------------|---------------------|---------------------|
| Distance   | 29             | Brain tumor resection (11); spinal surgery (8); ventriculostomy (6); transsphenoidal brain tumor resection (2); aneurysm repair (1); percutaneous rhizotomy (1) | 1390                | Most well-studied APM; used to represent accuracy (distance to target); used to represent efficiency (total distance traveled); distance-to-target metrics improved w/ increased experience; used in all machine learning algorithms studied |
| Time       | 25             | Brain tumor resection (12); spinal surgery (6); transsphenoidal brain tumor resection (2); aneurysm repair (2); ventriculostomy (3) | 1142                | Various uses: task completion, time of contact, time under fluoroscopy; few studies demonstrated repeated utility of time alone as an APM that distinguishes skill; shorter times could represent increased efficiency, but longer times could represent safety prioritization; complex, derived metrics that use time may be more valuable |
| Kinematics | 7              | Brain tumor resection (3); spinal surgery (2); transsphenoidal brain tumor resection (1); aneurysm repair (1) | 305                 | Encompasses tool movement data, such as velocity, acceleration, & jerk (first derivative of acceleration); different tasks & interventions lead to significantly different motion signatures; mixed evidence on utility of kinematics alone to differentiate surgical groups; kinematics consistently used by machine learning algorithms to predict training status |
| Force      | 19             | Brain tumor resection (12); spinal surgery (5); aneurysm repair (2) | 729                 | Early force metrics include max force applied, or cumulative sum of forces; more recent APMS have included “force histograms” or “force pyramids”; those w/ higher skill & improved task performance apply less force; force metrics commonly used by machine learning algorithms |
| Blood loss | 6              | Brain tumor resection (6)           | 162                 | High clinical relevance, but least-well-studied APM; top-performing surgeons experienced less blood loss; may be of increased relevance given difficulty of intraop blood loss measurement |
| Volume of resection | 13 | Brain tumor resection (10); spinal surgery (3) | 409 | Quantifies percentage of brain tumor & normal tissue removed; vol removed over time quantifies efficiency; experts tended to resect less tumor but had least amount of normal tissue removed |

Max = maximum.
APMs were divided by domain and then further subdivided by pathology. Total number of participants for each metric was recorded, and summary of analysis was outlined.
* Majority of studies incorporated multiple APM domains. Additionally, some studies used an overlapping set of participants with others; therefore, the values listed are upper bounds.
during VR brain tumor resection, and 1 was a VR hemilaminectomy model.7

A VR transphenoidal procedure to resect a pituitary tumor and an aneurysm clipping procedure used velocity, acceleration, and jerk as kinematic measures. Both found that experts made fewer regular movements, indicating that they are more effective with each motion without the need for repetition.18,32

Other studies have indirectly used kinematic performance metrics to understand motion patterns or gestures. Siyar et al. found less tremor, quantified by power spectral analysis at physiological frequencies between 8 and 12 Hz, in skilled (staff neurosurgeons and senior residents) compared to novices (junior residents and medical students).31 Their power spectral analysis relied primarily on instrument tip coordinates, instrument orientation, and force over time to identify tremor frequencies to differentiate groups.31

**Pressure, Force, and Haptic Feedback**

Force APMs, assessed through haptic systems when a tool comes into contact with tissue, include maximum force applied and summation of forces used. In addition, these metrics can be further subdivided based on left-versus right-handedness and the spatial distribution of the brain region in question. We found 19 studies that assessed force, 14 of which used the NeuroVR system (Supplemental Table 4).10-14,17,20,25,27,28,32-35,37-41 One NeuroTouch pilot study assessed the variability in psychomotor performance between a novice resident and expert neurosurgeon through novel force pyramids (a 3D representation of force along the plane of the tumor) and force histograms. These force metrics were able to differentiate between an expert neurosurgeon and a junior resident.11 Subsequent works compared force pyramids by handedness and hand ergonomics, as well as excess applied force, to detect microsurgical performance changes after sleep interruption.38-40

Maximum force applied during brain tumor resection on NeuroVR was 1 of 4 APMs identified by an artificial intelligence (AI) system to predict participants’ training levels.29

Three studies examined haptic feedback perception, which is similar to force measurement.14,26,27 Two of the 3 studies developed a scoring model composed of trajectory planning, motor planning, and haptic perception that successfully demonstrated that residents had improved sensory and motor performance when compared to medical students.25,26

**Virtual Surgical Outcome Measurements**

**Blood Loss**

Six studies measured volume of blood lost during a virtual brain tumor resection using the NeuroVR system (Supplemental Table 5).12,14,25,33,34,37 Experimental designs of included studies limit the assessment of blood loss as a VR-generated, reliable assessment metric. However, groups have developed several machine learning algorithms to differentiate surgical skill levels and included at least 1 blood loss APM in their final list of performance metrics.24,35

**Volume and Extent of Resection**

Thirteen articles used volume of resection as a metric for surgical performance (Supplemental Table 6).10-14,17,20,25,33-35,37,39 Eleven studies used the NeuroTouch/NeuroVR platform for brain tumor resection and hemilaminectomy, and 2 used the Sim-Ortho platform for anterior cervical discectomy and fusion.20,41

Extent of tumor resection is an important predictor of outcome in brain tumor patients and is a significant measure of the success of a resection. Studies used volume of resection to calculate extent of resection,10,40 determined the extent of tumor resection versus healthy tissue removal,20,40,41 and delineated the efficiency of surgical maneuvers (based on how much tumor is removed with each maneuver).41

As an APM, extent of resection could discriminate between resident and medical student performance, but not between residents at different levels of training.17 In a study comparing experts to trainees, expert neurosurgeons achieved a lower extent of resection compared to residents because experts prioritized preserving normal brain tissue.40

**Advanced Metrics, Machine Learning, and AI Models**

In addition to raw data, APMs were further modified to create advanced metrics. One advanced metric from NeuroTouch/NeuroVR is the efficiency index (the percentage of time spent actively resecting the brain tumor divided by total task time),12,14 and another was the development of force pyramids (a 3D representation of force along the plane of the tumor), both of which accurately distinguished surgical training levels.40

Four studies used machine learning and/or AI algorithms based on VR outputs to predict a participant’s training level group: NeuroVR (3 studies), Sim-Ortho (1 study).13,24,35,41 Studies used a variety of machine learning architectures, including K-nearest neighbor, naïve Bayes, support vector machines (SVMs), discriminant analysis, and conventional neural networks. Distance traveled by instruments, kinematics, force, bleeding, and volume of resection metrics taken from VR systems were all selected as inputs into various models. These models accurately predicted level of training, with a minimum accuracy of 65.9% and a maximum accuracy of 97.6% (Table 3) across studies.

**Discussion**

In our systematic review we identified studies that assessed surgeon performance in VR environments. We did not identify studies assessing surgeon performance in AR environments, which may present a future avenue for surgical education. We found that APMs based on kinematics, distance to target, force, blood loss, and volume of resection were available and used to assess surgeon performance. Surgeon performance was commonly defined by task success, classification of surgeons as expert or non-expert, and assessments of technical skill. The increasing number of publications in recent years (5 in the past year) demonstrates the growing interest in VR technology to assess neurosurgical performance.
of APMs, including kinematic APMs, assessments of he-
skill). NeuroVR has developed complex features to pro-
tools (e.g., objective structured assessment of technical
and other metrics that simulate known surgical assessment
mostasis, assessments of tumor/normal tissue interaction,
cutting, and removal.8


to recreate the physical world and maintain face and
construct validity. The latest generation of VR systems
have shown moderate realism with anatomy and overall
validity of tools (curette, Kerrison).20

The most obvious challenge of VR systems is their
need to recreate the physical world and maintain face and
construct validity. The latest generation of VR systems
have shown moderate realism with anatomy and overall
task difficulty; however, haptic feedback and realism of
tools continue to be rated poorly by users of many VR
systems.17,20 Compared to cadaveric simulations, VR sys-
tems do not fully recreate the operative environment.44–46
Whether this obstacle can be overcome or is an inherent
limitation to reconstructed fields is unknown.

Development of Performance Metrics With VR

The earliest VR models simulated a ventriculostomy and
pedicule screw placement and used largely rudimen-
tary APMs: distance to target and time for task comple-
tion.2,21–23 State-of-the-art VR systems now capture APMs
across various domains including kinematics, force, and
clinically relevant APMs like blood loss or volume of re-
section.12,35,41

These APMs have been used to predict surgeon perfor-
mance on tasks and to differentiate between surgical lev-
els of training. Even rudimentary APMs taken from early
VR systems could classify successful versus unsuccessful
EVD placements based on distance to the foramen of
Monro.2,20,16,47 Differences in surgical economy of motion
and tool instrument movement are routinely assessed via
objective structured assessment of technical skill and have
been taken from VR systems.35,45 VR may also be particu-
larly suited to evaluate metrics that are difficult to quantify
in a real clinical scenario; simulated blood loss and tumor
volume removed were more accurately measured through
the VR system than through visual assessment.35

However, other APMs may require further development

Comparison of VR Platforms

The NeuroVR (formerly NeuroTouch) platform has
been used to develop several neurosurgical modules, with
most users studying a brain tumor removal task. Other it-
erations of the platform have included hemilaminectomy
and aneurysm repair trials. For the brain tumor removal
task, participants hold a sucker in their left hand, an ul-
trasonic aspirator in their right hand, and interact with a
benchtop model head while viewing a VR screen (Fig. 2).8,44
The platform has developed to include a wide variety
of APMs, including kinematic APMs, assessments of he-
mostasis, assessments of tumor/normal tissue interaction,
and other metrics that simulate known surgical assessment
tools (e.g., objective structured assessment of technical
skill).42 NeuroVR has developed complex features to pro-
mote realism, including extensive haptic feedback (tool-
tool collisions, tool-tissue interaction), tissue deformation
and topology changes, and real-time simulation of rupture,
cutting, and removal.8

The ImmersiveTouch system was used in 8 studies for
hardware placement-type tasks including ventriculostomy
(5 studies), percutaneous rhizotomy, pedicule screw place-
ment, and spinal needle placement (1 study each). These
tasks assess accuracy of the surgeon’s placement/targeting
by using a high-definition 3D display and a haptic stylus to
allow for tool tracking and placement of the EVD, pedicle
screw, or other device. The ImmersiveTouch ventriculos-
tomy placement environment has high face validity and
incorporates tactile haptic feedback for the user. For ex-
ample, it simulates the “popping” sensation felt when pro-
gressing a catheter past the dense ependymal ventricular
lining and into the ventricle. However, it may not be as
realistic for other procedures.43

The final available product with repeated usage was the
Sim-Ortho VR simulator, in which users wore 3D glass-
es and received audio feedback while using virtual tools
along with an interactive screen.20 Sim-Ortho is the newest
of the VR technologies, and its interface measured force,
volume in voxels, path length, and other advanced APMs
similar to NeuroVR. Validity ratings were moderate, with
strengths in anatomical realism but weaknesses in validity
of tools (curette, Kerrison).20

The following key was used for the subdivision of participants under the assessment column: A = attending; F = fellow; J = junior resident; M = medical student; S = senior resident.

The APMs listed were collected by the most accurate machine learning algorithm for each study.

### TABLE 3. Studies that incorporated VR-derived metrics into a machine learning algorithm to predict training level of participants

| Authors & Year     | Assessment                  | Machine Learning Algorithms | Accuracy | Variables Incorporated | APMs Incorporated* |
|--------------------|-----------------------------|-----------------------------|----------|------------------------|---------------------|
| Mirchi et al., 202041 | Training level: A/F vs S vs J | Artificial neural network   | 83.3%    | 13                     | Distance; time; kinematics; force; blood loss; volume resected; contacts made w/ tissue; angle of instrument |
| Winkler-Schwartz et al., 201935 | Training level: A vs F/S vs J vs M | K-nearest neighbor Naive Bayes Discriminant analysis SVM | 90%       | 6                      | Distance; kinematics; force |
| Bissonnette et al., 201913 | Training level: A/F/S vs J/M | SVM K-nearest neighbor Linear discriminant analysis Decision tree analysis Naive Bayes | 97.6%     | 12                    | Distance; time; kinematics; force; contacts made w/ tissue; angle of instrument |
| Mirchi et al., 202025 | Training level: A/F/S vs J/M | SVM | 92% | 4 | Distance; kinematics; force; blood loss |

8 The final available product with repeated usage was the Sim-Ortho VR simulator, in which users wore 3D glasses and received audio feedback while using virtual tools along with an interactive screen.20 Sim-Ortho is the newest of the VR technologies, and its interface measured force, volume in voxels, path length, and other advanced APMs similar to NeuroVR. Validity ratings were moderate, with strengths in anatomical realism but weaknesses in validity of tools (curette, Kerrison).20

The most obvious challenge of VR systems is their need to recreate the physical world and maintain face and construct validity. The latest generation of VR systems has shown moderate realism with anatomy and overall task difficulty; however, haptic feedback and realism of tools continue to be rated poorly by users of many VR systems.17,20 Compared to cadaveric simulations, VR systems do not fully recreate the operative environment.44–46 Whether this obstacle can be overcome or is an inherent limitation to reconstructed fields is unknown.
to accurately assess surgical skill. Force and volume metrics were not able to differentiate practicing neurosurgeons from trainees. Similarly, Teodoro-Vite et al. found no difference in kinematic movements between staff neurosurgeons and trainees in an aneurysm clipping task. Nonetheless, whereas these APMs as raw data points may be unable to differentiate surgeon performance, transformations or second-order APMs, such as “force pyramids,” were able to distinguish between classes of surgeons with significant accuracy.

In addition to second-order metrics, machine learning and AI models have been developed that use APMs to classify participant training level. These models have incorporated kinematic data, positional data, and force data into their models to develop accurate predictions. Given that surgical metrics are largely interrelated, their use in a combined model may provide the most accurate metrics for surgical assessment and characterization. The clinical utility of these models and their validity across different tasks and training programs remain open questions, but these approaches point to a promising direction in the assessment of surgeon performance.

More recently, VR has been used as a test bed to evaluate the effects of interventions on surgeon performance, including studies of sleep disruption or acute stress. However, the results of these experiments have often contradicted hypotheses (e.g., sleep deprivation improving performance), indicating a need for further validity assessments. Nonetheless, the virtual environment provides a safe, low-stakes method to assess changes in surgeon performance. VR is a promising tool that may help organizations target interventions to improve patient safety and outcomes.

External Validation

In other industries, such as aviation, VR environments serve as an important training tool and low-stakes practice arena for the routine assessment and continued training of pilots. Within neurosurgery, however, VR is currently an adjunctive tool for surgical education and is not routinely incorporated in the assessment of surgical skill progression. Further external validation of these systems is needed to bridge this gap. Current studies have not evaluated VR-derived metric progression throughout a residency program or demonstrated that differences in VR system performance mirror true intraoperative performance. Accordingly, questions regarding frequency of assessment, transferability to the operating room, and the logistical application of VR systems into a residency curriculum are largely unanswered.

Augmented Reality

Our initial aim was to identify both VR and AR systems that provide automatic, quantifiable metrics to assess surgeon performance. Whereas these metrics are well described within VR settings, we found no articles describing the use of AR in surgeon performance assessment. Despite positive user ratings and perceived transferability of VR simulations, VR is inherently removed from a true clinical setting. AR offers the greatest potential of transfer to the operating room given that an AR layer could be added to users manipulating real surgical instruments and performing cadaveric dissection tasks. AR systems can gather data by using video analysis and can display data on top of real tissue or in real procedures. AR in its current form is primarily used for preoperative planning, neuronavigation, or stereotaxis. Although AR displays are used to show stereotactic information such as pedicle screw trajectories, current AR systems do not dynamically interact with the surgeon and the environment, and any performance metric is derived from imaging studies evaluating the outcome of the AR system. This is a serious limitation to the current status of AR for surgeon performance improvement. AR systems could display relevant case information, metrics, distances, force measurements, etc., that would be automatically collected and analyzed, which would provide peerless assessments of technical skill in a true operative setting.

Limitations

Neurosurgical performance assessment in which VR is used is a field in its infancy. Many of the VR systems studies were performed by a small number of groups using the same software, potentially biasing the true ability of VR software as a whole. There may also be significant publication bias, because papers that show large differences in groups would be more likely to be published than papers demonstrating that metrics showed no differences. Given that many VR systems are developed by industry, this may introduce conflicts into published studies. This systematic review is limited—as all are—by availability of currently published studies, variability in study designs, and limited sample sizes within each trial. Additionally, our ability to draw conclusions is limited by the institutions that have tested VR systems, and their successes and failures may not be representative of their utility to neurosurgery training programs around the world.

Conclusions

After 30 years of evolution, VR environments are now able to provide meaningful feedback to surgeons and are improving on an annual basis. VR-generated APMs, such as distance, time, force, and volume, can differentiate levels of training of surgeons and track trainee progress through a curriculum. With further development and validation, future tools might be developed to measure surgeon competency and to teach new procedures. External validation and longitudinal tracking of VR-trained surgeon outcomes would be welcome developments in this field. Unlike VR, currently available AR systems do not provide meaningful surgeon performance assessment. We anticipate that future generations of AR systems may overcome these limitations. Both VR and AR systems should be considered for integration into standard neurosurgical training program curricula.

References

1. Bernardo A. Virtual reality and simulation in neurosurgical training. World Neurosurg. 2017;106:1015-1029.
2. Banerjee PP, Luciano CJ, Lemole GM Jr, Charbel FT, Oh MY. Accuracy of ventriculostomy catheter placement using a head- and hand-tracked high-resolution virtual reality simulator with haptic feedback. J Neurosurg. 2007;107(3):515-521.
3. Clarke DB, D’Arcy RCN, Delorme S, Laroche D, Godin G, Hajra SG, et al. Virtual reality simulator: demonstrated use in neurosurgical oncology. Surg Innov. 2013;20(2):190-197.

4. Fiani B, De Stefano F, Kondili A, Covarrubias C, Reier L, Sarhadi K. Virtual reality in neurosurgery: “Can you see it?” — A review of the current applications and future potential. World Neurosurg. 2020;141:291-298.

5. Meola A, Cutolo F, Carbone M, Cagnozzo F, Ferrari M, Ferrari V. Augmented reality in neurosurgery: a systematic review. Neurosurg Rev. 2017;40(4):537-548.

6. Mitha AP, Almekhlafi MA, Janjua MJJ, Albuquerque FC, McDougall CG. Simulation and augmented reality in endo-vascular neurosurgery: lessons from aviation. Neurosurgery. 2013;72(suppl 1):107-114.

7. Tagaytayan R, Kelemen A, Sik-Lanyi C. Augmented reality in neurosurgery. Arch Med Sci. 2018;14(3):572-578.

8. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. Neurosurgery. 2012;71(1 Suppl Operative):32-42.

9. Abhari K, Baxter JSH, Chen ECS, Khan AR, Peters TM, de Ribauipierre S, Eagleson R. Training for planning tumour resection: augmented reality and human factors. IEEE Trans Biomed Eng. 2015;62(6):1466-1477.

10. AlZhrani G, Alotaibi F, Azarnoush H, Winkler-Schwartz A, Sabbagh A, Bajunaid K, et al. Proficiency performance benchmarks for removal of simulated brain tumors using a virtual reality simulator NeuroTouch. J Surg Educ. 2015;72(4):685-696.

11. Azarnoush H, AlZhrani G, Winkler-Schwartz A, Alotaibi F, Gelinas-Phaneuf N, Pazos V, et al. Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. Int J CARs. 2015;10(5):603-618.

12. Bajunaid K, Mullah MAS, Winkler-Schwartz A, Alotaibi FE, Fares J, Baggiani M, et al. Impact of acute stress on psychomotor bimanual performance during a simulated tumor resection task. J Neurosurg. 2017;126(1):71-80.

13. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF. Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. J Bone Joint Surg Am. 2019;101(23):e127.

14. Bugdaji A, Sowaya R, Bajunaid K, Olvi D, Winkler-Schwartz A, Ledwos N, et al. Is virtual reality surgical performance influenced by force feedback device utilized? J Surg Educ. 2019;76(1):262-273.

15. Burström G, Nachabe R, Persson O, Edström E, Elmi Terand-A. Augmented and virtual reality instrument tracking for minimally invasive spine surgery: a feasibility and accuracy study. Spine (Phila Pa 1976). 2019;44(15):1097-1104.

16. Cagiltay NE, Ozcelik E, Sengul G, Berker M. Construct and face validity of the educational computer-based environment (ECE) assessment scenarios for basic endoneurosurgery skills. Surg Endosc. 2017;31(11):4485-4495.

17. Gelinas-Phaneuf N, Choudhury N, Al-Habib AR, Cabral A, Nadeau E, Mora V, et al. Assessing performance in brain tumor resection using a novel virtual reality simulator. Int J Comput Assist Radiol Surg. 2014;9(1):1-9.

18. Heredia-Pérez SA, Harada K, Padilla-Castañeda MA, Marques-Marinho M, Márquez-Flores JA, Mitsuishi M. Virtual reality simulation of robotic transphenoidal brain tumor resection: evaluating dynamic motion scaling in a master-slave system. Int J Med Robot. 2019;15(1):e1953.

19. Hooten KG, Lister JR, Lombard G, Lizdas DE, Lampotang S, Rajon DA, et al. Mixed reality ventriculostomy simulation: experience in neurosurgical residency. Neurosurgery. 2014;10(suppl 4):576-581.

20. Ledwos N, Mirchi N, Bissonnette V, Winkler-Schwartz A, Yilmaz R, Del Maestro RF. Virtual reality anterior cervical discectomy and fusion simulation on the novel Sim-Ortho platform: validation studies. Oper Neurosurg (Hagerstown). 2020;20(1):74-82.

21. Lemole GM Jr, Banerjee PP, Luciano C, Neckrysh S, Charbel FT. Virtual reality in neurosurgical education: part-task ventriculostomy simulation with dynamic visual and haptic feedback. Neurosurgery. 2007;61(1):142-149.

22. Luciano CJ, Banerjee PP, Bellotte B, Oh GM, Lemole M Jr, Charbel FT, Roitberg B. Learning retention of thoracic pedicle screw placement using a high-resolution augmented reality simulator with haptic feedback. Neurosurgery. 2011;69(1 Suppl Operative):ons14-ons19.

23. Luciano CJ, Banerjee PP, Sorenson JM, Foley KT, Ansari SA, Rizzi S, et al. Percutaneous spinal fixation simulation with virtual reality and haptics. Neurosurgery. 2015;72(suppl 1):89-96.

24. Mirchi N, Bissonnette V, Ledwos N, Winkler-Schwartz A, Yilmaz R, Karlrik B, Del Maestro RF. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. Oper Neurosurg (Hagerstown). 2020;19(1):65-75.

25. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The Virtual Operative Assistant: an explainable artificial intelligence tool for simulation-based training in surgery and medicine. PLoS One. 2020;15(2):e0229596.

26. Perin A, Galbiati TF, Gambatesa E, Ayadi R, Orena EF, Cuomo V, et al. Filling the gap between the OR and virtual simulation: a European study on a basic neurosurgical procedure. Acta Neurochir (Wien). 2018;160(11):2087-2097.

27. Roitberg B, Banerjee P, Luciano C, Matulayuskas M, Rizzi S, Kania P, Gasco J. Sensory and motor skill testing in neurosurgery applicants: a pilot study using a virtual reality haptic neurosurgical simulator. Neurosurgery. 2013;73(suppl 1):116-121.

28. Roitberg BZ, Kania P, Luciano C, Dharmavaram N, Banerjee P. Evaluation of sensory and motor skills in neurosurgery applicants using a virtual reality neurosurgical simulator: the sensory-motor quotient. J Surg Educ. 2015;72(6):1165-1171.

29. Schirmer CM, Elder JB, Roitberg B, Lobel DA. Virtual reality-based simulation training for ventriculostomy: an evidence-based approach. Neurosurgery. 2013;73(suppl 1):66-73.

30. Shakur SF, Luciano CJ, Kania P, Roitberg BZ, Banerjee PP, Slavin KV, et al. Usefulness of a virtual reality percutaneous trigeminal rhizotomy simulator in neurosurgical training. Neurosurgery. 2015;71(suppl 3):420-425.

31. Siyar S, Azarnoush H, Rashiidi S, Del Maestro RF. Tremor assessment during virtual reality brain tumor resection. J Surg Educ. 2020;77(3):643-651.

32. Teodoro-Vite S, Pérez-Lomelí JS, Domínguez-Velasco CF, Hernández-Valencia AF, Capurso-García MA, Padilla-Castañeda MA. A high-fidelity hybrid virtual reality simulator of aneurysm clipping repair with brain sylvian fissure exploration for vascular neurosurgery training. J Surg Educ. 2020;77(3):643-651.

33. Teodoro-Vite S, Pérez-Lomelí JS, Domínguez-Velasco CF, Hernández-Valencia AF, Capurso-García MA, Padilla-Castañeda MA. A high-fidelity hybrid virtual reality simulator of aneurysm clipping repair with brain sylvian fissure exploration for vascular neurosurgery training. J Surg Educ. 2020;77(3):643-651.

34. Winkler-Schwartz A, Lozano MAS, Marwa I, Alotaibi FE, Fares J, et al. Bimanual psychomotor performance in neurosurgical resident applicants assessed using NeuroTouch, a virtual reality simulator. J Surg Educ. 2016;73(6):942-953.

35. Winkler-Schwartz A, Marwa I, Bajunaid K, Mullah M, Alotaibi FE, Bugdaji A, et al. Comparison of visual rating scales and simulated virtual reality metrics in neurosurgical training: a generalizability theory study. World Neurosurg. 2019;127:65-75.
pertise in virtual reality simulation. *JAMA Netw Open*. 2019; 2(8):e198363.

36. Yudkowsky R, Luciano C, Banerjee P, Schwartz A, Alaraj A, Lemole GM Jr, et al. Practice on an augmented reality/haptic simulator and library of virtual brains improves residents’ ability to perform a ventriculostomy. *Simul Healthc*. 2013; 8(1):25-31.

37. Holloway T, Lorsch ZS, Chary MA, Sobotka S, Moore MM, Costa AB, et al. Operator experience determines performance in a simulated computer-based brain tumor resection task. *Int J CARS*. 2015;10(1):1853-1862.

38. Micko A, Knopp K, Knosp E, Wolfsberger S. Microsurgical performance after sleep interruption: a neurotouch simulator study. *World Neurosurg.* 2017;106:92-101.

39. Sawaya R, Alsideiri G, Bugardi A, Winkler-Schwartz A, Azarnoush H, Bajunaid K, et al. Development of a performance model for virtual reality tumor resections. *J Neurosurg*. 2018; 131(1):192-200.

40. Azarnoush H, Siai S, Sawaya R, Zhraani GA, Winkler-Schwartz A, Alotaibi FE, et al. The force pyramid: a spatial analysis of force application during virtual reality brain tumor resection. *J Neurosurg*. 2017;127(1):171-181.

41. Mirchi N, Bissonnette V, Ledwos N, Winkler-Schwartz A, Yilmaz R, Karlik B, Del Maestro RF. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. *Oper Neurosurg (Hagerstown)*. 2020;19(1):65-75.

42. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.

43. Malone HR, Syed ON, Downes MS, D’Ambrosio AL, Quest DO, Kaiser MG. Simulation in neurosurgery: a review of computer-based simulation environments and their surgical applications. *Neurosurgery*. 2010;67(4):1105-1116.

44. Baksheshian J, Strickland BA, Patel NN, Jakoi AM, Minneti M, Zada G, et al. The use of a novel perfusion-based cadaveric simulation model with cerebrospinal fluid reconstitution comparing dural repair techniques: a pilot study. *Spine J*. 2017;17(9):1335-1341.

45. Carey JN, Minneti M, Leland HA, Demetriades D, Talving P. Perfused fresh cadavers: method for application to surgical simulation. *Am J Surg*. 2015;210(1):179-187.

46. Donoho DA, Johnson CE, Hur KT, Buchanan 1A, Fredrickson VL, Minneti M, et al. Costs and training results of an objectively validated cadaveric perfusion-based internal carotid artery injury simulation during endoscopic skull base surgery. *Int Forum Allergy Rhinol*. 2019;9(7):787-794.

47. Lemole M, Banerjee PP, Luciano C, Charbel F, Oh M. Virtual ventriculostomy with ‘shifted ventricle’: neurosurgery resident surgical skill assessment using a high-fidelity haptic/graphic virtual reality simulator. *Neurrol Res*. 2009;31(4):430-431.

48. Landman A, van Oorschot P, van Paassen MMR, Groen EL, Bronkhorst AW, Mulder M. Training pilots for unexpected events. A simulator study on the advantage of unpredictable and variable scenarios. *Hum Factors*. 2018;60(6):793–805.

49. Molina CA, Scuibba DM, Greenberg JK, Khan M, Witham T. Clinical accuracy, technical precision, and workflow of the first in human use of an augmented-reality head-mounted display stereotactic navigation system for spine surgery. *Oper Neurosurg (Hagerstown)*. 2021;20(3):300-309.

50. Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. *Clin J Am Soc Nephrol*. 2008;3(1):253-260.

Disclosures
Dr. Hung is a consultant for Johnson & Johnson, Quantgene, and Mimic Technologies.

Author Contributions
Conception and design: Chan, Pangal. Acquisition of data: Chan, Pangal. Analysis and interpretation of data: Chan, Pangal. Drafting the article: Chan, Pangal. Cardinal. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Zada. Administrative/technical/material support: Zada, Chan, Pangal, Cardinal, Donoho. Study supervision: Zada, Donoho.

Supplemental Information
Online-Only Content
Supplemental material is available online.
*Supplemental Tables 1–6*. https://thejns.org/doi/suppl/10.3171/2021.5.FOCUS21210.

Correspondence
Gabriel Zada: Keck School of Medicine of the University of Southern California, Los Angeles, CA. gabriel.zada@med.usc.edu.