Construct Validity of the Endoscopic Sinus Surgery Simulator

II. Assessment of Discriminant Validity and Expert Benchmarking

Marvin P. Fried, MD; Babak Sadoughi, MD; Suzanne J. Weghorst, MS, MA; Michael Zeltsan, MS; Hernando Cuellar, MD; Jose I. Uribe, MD; Clarence T. Sasaki, MD; Douglas A. Ross, MD; Joseph B. Jacobs, MD; Richard A. Lebowitz, MD; Richard M. Satava, MD

Objectives: To establish discriminant validity of the endoscopic sinus surgery simulator (ES3) (Lockheed Martin, Akron, Ohio) between various health care provider experience levels and to define benchmarking criteria for skills assessment.

Design: Prospective multi-institutional comparison study.

Setting: University-based tertiary care institution.

Participants: Ten expert otolaryngologists, 14 otolaryngology residents, and 10 medical students.

Interventions: Subjects completed the ES3’s virtual reality curriculum (10 novice mode, 10 intermediate mode, and 3 advanced mode trials). Performance scores were recorded on each trial. Performance differences were analyzed using analysis of variance for repeated measures (experience level as between-subjects factor).

Main Outcome Measures: Simulator performance scores, accuracy, time to completion, and hazard disruption.

Results: The novice mode accurately distinguished the 3 groups, particularly at the onset of training (mean scores: senior otolaryngologists, 66.0; residents, 42.7; students, 18.3; for the paired comparisons between groups 1 and 2 and groups 1 and 3, \( P = .04 \) and \( .03 \), respectively). Subjects were not distinguished beyond trial 5. The intermediate mode only discriminated students from other subjects \( (P = .008) \). The advanced mode did not show performance differences between groups. Scores on the novice mode predicted those on the intermediate mode, which predicted advanced mode scores \( (r = 0.687) \), but no relationship was found between novice and advanced scores. All groups performed equally well and with comparable consistency at the outset of training. Expert scores were used to define benchmark criteria of optimal performance.

Conclusions: This study completes the construct validity assessment of the ES3 by demonstrating its discriminant capabilities. It establishes expert surgeon benchmark performance criteria and shows that the ES3 can train novice subjects to attain those. The refined analysis of trial performance scores could serve educational and skills assessment purposes. Current studies are evaluating the transfer of surgical skills acquired on the ES3 to the operating room (predictive validity).

Arch Otolaryngol Head Neck Surg. 2007;133:350-357

Surgical residency programs rarely, if ever, evaluate the manual dexterity of medical student applicants as a criterion for admission. Likewise, the innate abilities of residents are not assessed at the outset of training, and it is hoped that they are acquired during the residency years. Being able to reliably measure trainee skills and progress is critical to residency program directors.

Although it has been estimated that judgment comprises 75% of a successful operation and technical skill, 25%, this latter factor has traditionally been overlooked in the evaluation of surgical trainees. Objective measures are rarely applied to surgical tasks, and the acquisition of skills is infrequently done with inanimate objects or devices, requiring experience with live patients to master the tasks. These concerns have a direct impact on the quality and safety of patient care. Martin et al have developed and validated the Objective Structured Assessment of Technical Skill for Surgical Residents examination to objectively assess laparoscopic surgical technical skill. The examination consists of six 15-minute tasks that evaluate basic surgical skills, such as bowel anastomosis or T-tube insertion, and may be performed either on live animals or at bench top stations with similar results. Both task-specific checklists and global rating scales were shown to be reliable and valid in the evaluation of residents be-
cause there was high interrater agreement and scores correlated with level of experience. Training on the bench stations was also shown to transfer well to performing similar procedures on a cadaver, suggesting that the skills developed in the simulation laboratory may transfer to the operating room (OR).

Other investigators have attempted to develop mechanical or computerized techniques to objectively measure surgical skill. Intraoperative motion analysis of surgical tools using electromagnetic trackers has provided measures of surgical skill that are more accurate than simple time-to-completion observations or subjective evaluation of performance. Performance on an advanced endoscopic psychomotor tester in the training box environment correlates well with subjective evaluation of operative skill and may also serve as an aptitude test on surgical ability. Such approaches help respond to the need for objective evaluation of surgical skill. In the near future, virtual reality (VR) may provide an even more thorough and malleable mode of surgical evaluation.

Currently, the training curriculum of otolaryngology residents in endoscopic sinus surgery (ESS) includes video material, cadaver dissections, and direct observation in the OR, where most of the training takes place. As residents progress under supervision by an experienced surgeon, they are given more of an active role in the procedure, ultimately becoming the major participant by their final year of training. However, once a procedure is completed, whether on cadaver or live subjects, repetition is made impossible by permanent alterations, and procedures may not be “erased” and restarted.

In response to the need for a simulator to train novice sinus surgeons, Department of Defense contractor Lockheed Martin, Inc (Akron, Ohio) gathered a team including the University of Washington Human Interface Technology Laboratory (Seattle), the Ohio Supercomputer Center (Columbus), and Immersion Corp (San Jose, Calif) to develop a training device, the ESS simulator (ES3) using both visual and haptic (force) feedback in a VR environment. The Lockheed Martin simulation development team built a physical space with embedded VR elements that allows for analysis of single or multiple tasks and their relationships. Unlike real patient or cadaver surgery, metrics can be applied to these experiences, and procedures can be repeated on demand.

The ES3 is a procedural simulator that trains and assesses the performance of an entire task, such as injection, which requires navigation, ambidexterity, and accuracy. The hardware comprises 4 principal components (Figure 1): (1) an SGI Octane workstation (Silicon Graphics Inc, Mountain View, Calif), which serves as the simulation host platform; (2) a personal computer–based haptic controller, providing control and coordination between a universal physical instrument handle and a set of virtual surgical instruments; (3) a personal computer–based voice recognition–enabled instructor, which operates the simulator by responding to spoken commands; and (4) an electromechanical platform, which serves as the human interaction interface with a replica of an endoscope, a surgical tool handle, and a rubber–headed mannequin.

The ES3 automatically collects performance data and analyzes them in real time to provide an accurate display of performance errors for proximate feedback and optimal skill acquisition. The ES3 also archives data for end results analysis and reporting.

Although simulation has historically been central to aviation safety, and high-fidelity VR trainers have had a longstanding impact in measurably improving the skill level of military and commercial pilots, simulators are still at the outset of their development in medicine, holding similar promises for the field.

The potential benefits of VR surgical simulators are numerous: objective assessment of surgical skill, elimination of patient risk as surgeons progress along their learning curves in the setting of VR rather than in actual surgery, simulation of complex procedures, standardization of residency training regardless of environment, and the ability to rehearse a procedure on patient–specific anatomy.

To date, a large variety of VR surgical trainers have been developed, ranging from a simple simulation of laparoscopic drills running on a personal computer to an advanced hepatic surgery simulator running on a high–end graphics workstation. Fields such as neurological surgery, ophthalmic surgery, laparoscopic surgery, open abdominal surgery, anastomosis, knee and shoulder arthroscopy, bronchoscopy, and intravenous catheter insertion already have been working on the development of simulation devices. Experts anticipate that by 2010, surgical simulators will have reached a level of performance where they can be acceptable as tools for testing and certification.

The potential to optimize surgical training and thereby diminish errors and risks to the patient by the use of VR surgical simulation environments is evident. The attrac-
tenness of computer-generated surgical simulation is its ability to create an environment that emulates the real surgical world without the inherent risks to patients.

Surgical simulators must demonstrate that they are both instructionally effective and valid for the evaluation of surgical skills. One fundamental and probably overarching component of simulator validity is construct validity. Defined as a “set of procedures for evaluating a testing instrument based on the degree to which test items identify the quality, ability, or trait it was designed to measure,”11 (p1526) construct validity assessment may be broken down into 2 main subcategories: convergent validity and discriminant validity. Defined as the agreement among measures that theoretically should be related, and discriminant validity, defined as the actual relationship between measures that theoretically should not be related.12

Our prior work on the construct validity of the ES3 showed a significant correlation between performance on the simulator and scores obtained on instruments for visuospatial and psychomotor skills evaluation, thus supporting the simulator’s convergent validity component.13 The study presented herein will complete these observations and finalize the assessment process of the ES3’s construct validity by presenting further work to study its discriminant validity component. In particular, the ability of the ES3 to differentiate between populations with various levels of surgical experience was evaluated and the possibility of using expert-level data as benchmarking criterion for skills assessment discussed. There is a strong need for a standardized format based on objective criteria for the measurement and analysis of technical skills. In the future, these standards will certainly include comparison of data from surgical simulator prototypes with measurements of physical variables.

### METHODS

#### SUBJECTS

The subjects enrolled in this study comprised senior otolaryngologists, otolaryngology residents, and medical students from various major teaching institutions in the Northeast. All subjects were contacted by telephone and electronic mail. Those who agreed to participate and who met the inclusion criteria were enrolled. Institutional review board clearance was obtained at each institution prior to data collection.

Ten practicing otolaryngologists, proficient in ESS (ie, who had performed more than 200 unsupervised ESS procedures outside of residency training and who were currently performing a minimum of 1 or 2 ESS procedures per week), 14 otolaryngology residents (who had performed fewer than 5 ESS procedures), and 10 medical students (who had no experience of ESS) were given a curriculum with reading and multimedia material, demonstration videos, selected publications,14,15 and a demonstration of the simulator. Familiarity with this curriculum, which was ensured by allowing participants 4 weeks to review it, was a prerequisite for training on the ES3. None of the participating subjects had previous experience with the ES3.

#### PROTOCOL

All subjects were asked to perform a total of 23 trials on the ES3: 10 trials on the novice mode, followed by 10 trials on the intermediate mode and 3 trials on the advanced mode. The novice mode is an abstract environment for basic skills practice, sequentially displaying a series of custom-made geometrical objects designed to teach the use of most commonly available ESS instruments. The tasks performed on the ES3’s surgical modes (intermediate and advanced) are described in the Table. All trials yielded a total score ranging from 0 (poorest performance) to 100 (best possible performance).

#### STATISTICAL ANALYSIS

After replacing missing trial values by series mean projections as per recommended statistical approaches,10,11 selective differences between the scores of each group were analyzed using 1-way analysis of variance. Dynamic differences between trial series of each group were analyzed using analysis of variance for repeated measures with experience-level group as the between-subjects factor. Equality of variances was verified by the Mauchly test of sphericity (significance level, P<.05), and violations of sphericity corrected by a Greenhouse-Geisser estimate. Specific contrasts were obtained with Scheffe and/or Bonferroni post hoc F tests. Correlations were assessed through Pearson product moment correlation coefficient. All data were computed and analyzed using SPSS statistical software for Windows (release 14.0.1; SPSS Inc, Chicago, Ill).

#### RESULTS

Enrolled subjects were assigned to 1 of 3 groups according to their level of experience at the time of inclusion: senior otolaryngologists (n=10), otolaryngology residents (n=14), and medical students (n=10). All subjects completed the full ES3 curriculum except for 2 students who dropped out of the study after completing the intermediate mode. The total number of trials performed by all subjects across all groups was 776, of which 11 yielded corrupt data caused by system failures or other unexpected events not related to the study protocol and requiring aborting some trial sessions. Those 11 sets of scores were considered as missing values and managed as described in the previous subsection. The results were analyzed according to simulator modes (novice, intermediate, advanced).

#### NOVICE MODE

Mean total scores are represented in Figure 2. The 3 groups started their training at different score levels,
with performance of senior otolaryngologists (mean score, 66.0) being superior to that of residents (mean score, 42.7; \( P = .04 \)), which was in turn superior to that of students (mean score, 18.3; \( P = .03 \)). This trend is consistent until at least trial 5. From trial 6 on, the 3 groups achieved similar scores and cannot be distinguished. These observations are confirmed by comparison of 2 subsets of trials: on trial series 1 through 5, there was a statistically significant difference between groups (\( F_2 = 11.7; P < .001 \)), and contrast is observed on all paired comparisons (students vs residents, \( P = .04 \); residents vs senior otolaryngologists, \( P = .04 \); students vs senior otolaryngologists, \( P < .001 \)); on analysis of trial series 6 through 10, performance of all groups was comparable (\( F_2 = 0.7; P = .53 \)), and no specific contrast was established (students vs residents, \( P = .91 \); residents vs senior otolaryngologists, \( P = .74 \); students vs senior otolaryngologists, \( P = .54 \)).

**Figure 2** displays the gradual progression of mean scores with their 95% confidence intervals (CIs) throughout the 10 trials. **Figure 4** summarizes the findings between trials 1 and 10. On trial 1, the mean student score was 18.3 (95% CI, 10.0-35.6), the mean resident score was 42.7 (95% CI, 28.9-56.6), and the mean senior otolaryngologist score was 66.0 (95% CI, 57.8-74.1). Between-group differences are significant (\( F_2 = 12.7; P < .001 \)) and verified on all paired comparisons (students vs residents, \( P = .03 \); residents vs senior otolaryngologists, \( P = .04 \); students vs senior otolaryngologists, \( P < .001 \)). By trial 10, all subjects improved significantly and brought their scores within a narrow range with remarkably decreased variabilities: mean student score, 89.7 (95% CI, 83.5-95.8); mean resident score, 89.6 (95% CI, 85.2-94.1); and mean senior otolaryngologist score, 91.1 (95% CI, 88.0-94.2). No significant difference was found between groups at this point of the training (\( F_2 = 0.15; P = .87 \)).

**INTERMEDIATE MODE**

Total scores analysis was not able to distinguish any group at the intermediate training level of the ES3. The study had to be narrowed down to individual analysis of dissection times to establish a distinction only in the performance of students compared with residents and senior otolaryngologists, with a cutoff identified this time after trial 4 (**Figure 5**). From trials 1 through 4, dissection times show a difference in performance according to groups (\( F_2 = 5.6; P = .008 \)), and specific contrasts show that the observed difference lies between students and both other groups (students vs residents, \( P = .01 \); students vs senior otolaryngologists, \( P = .04 \)), whereas residents and senior otolaryngologists achieved similar performance (\( P = .95 \)). From trial 5 on, the 3 groups achieved similar levels of performance (\( F_2 = 1.9; P = .17 \)); therefore, no specific contrast is noted on paired compari-
sons (students vs residents, \(P = .18\); residents vs senior otolaryngologists, \(P = .89\); students vs senior otolaryngologists, \(P = .43\)).

The progression of dissection times between trials 1 and 10 is shown in Figure 6. Students started at a mean time of 1488 seconds (95% CI, 919-2056), which is significantly higher (\(F_2 = 4.4\); \(P = .02\)) than the mean time achieved by residents (886 seconds [95% CI, 756-1017]; \(P = .02\)) and senior otolaryngologists (1021 seconds [95% CI, 736-1309]). The mean times of residents and senior otolaryngologists at trial 1 have no statistically significant difference (\(P = .81\)). Again, at trial 10, all groups performed comparably well (\(F_2 = 1.1\); \(P = .35\)) and markedly decreased their ranges of performance variation (mean student time, 484 seconds [95% CI, 296-671]; mean resident time, 395 seconds [95% CI, 355-434]; mean senior otolaryngologist time, 462 seconds [95% CI, 398-521]).

**ADVANCED MODE**

The data gathered at the advanced level of training (3 trials) were not able to distinguish patterns specific to any group and did not seem useful for discrimination purposes. Even refined analysis of dissection times did not reflect any statistically significant difference between groups, whether at trial 1 of the advanced mode (\(F_2 = 0.8\); \(P = .47\)) or at trial 3 (\(F_2 = 3.0\); \(P = .06\)), which was the last session of the entire VR training provided by the ES3.

**CROSS-COMPARISONS AND CORRELATIONS**

Mean individual scores obtained across all subject groups and trial sessions were compared among various ES3 training modes, and longitudinal correlations were established by performance of individuals, regardless of their level of experience. A significant positive score correlation was observed between the novice and intermediate modes of the ES3 (\(r = 0.453\); \(P = .007\)) and an even stronger positive score correlation observed between the intermediate and advanced modes (\(r = 0.687\); \(P < .001\)). However, no transitivity relationship was firmly established because no statistically significant direct correlation was found between scores obtained on the novice and advanced modes (\(r = 0.15\); \(P = .39\)).

**BENCHMARKING DATA**

Scores obtained by senior otolaryngologists were retrieved to establish benchmark criteria of expert-level performance. This process was limited to trials 7 through 10 on the intermediate mode to take into consideration only those trials in which a final plateau of performance was consistently maintained. The resulting mean (SD) overall score observed was 93.9 (9.8).

**COMMENT**

These results clearly demonstrate that the ES3’s novice mode is able to distinguish 3 levels of experience, although the learning curves from the 3 groups improve asymptotically until they gather in a quite narrow corridor of high performance. This means that the training provided leads to the desired level of performance within the defined number of trials and that the generated skill is consistent regardless of the initial level of experience or baseline variability of performance. The latter observations apply to the intermediate mode as well, although the same discrimination capabilities were not found here. Students were distinguished from other subjects, but the intermediate mode did not reflect any difference between residents and senior otolaryngologists. As for the advanced mode, it neither showed significant
improvement in performance with trials nor was able to decipher subject group assignments.

Our experience suggests that the decrease in the power of discrimination as the trainer's difficulty level increases may be explained by the fact that each level provides adequate preparedness for the next. This is verified by the correlations observed between scores on different modes. However, the lack of good correlation between the novice mode and advanced mode scores was quite unexpected. Our best explanation of this observation is that the skill set taught on the novice mode is useful in the completion of the intermediate mode, and the skill set taught on the intermediate mode is in turn useful in the completion of the advanced mode, but there is not necessarily a significant overlap between the two. In other words, completing the novice mode is not sufficient to perform well on the advanced mode because many of the assessed tasks are presented only in the intermediate mode. Furthermore, the resemblance between the intermediate and advanced modes is certainly much greater than that between the novice and intermediate modes, which is probably why all groups concur at a similar plateau as early as at trial 1 of the advanced mode.

Another explanation for the correlation discrepancies may be prior intrinsic differences between subjects. Again, the skill sets required are different depending on each ES3 mode, and one may hypothesize that psychomotor and visuospatial abilities are preferentially rated by the abstract environment of the novice mode, whereas the surgical environment of the intermediate and advanced modes summon other types of dexterity more directly linked to the skills expected from a surgeon. Therefore, there might be a variable process of skills selection throughout the ES3 curriculum. Interestingly, medical students start out with subpar performances and improve over time but take longer than residents to reach the level of the expert surgeons for the parameters studied, particularly in the intermediate mode. This may reflect the fact that, in addition to technical skills, a comprehensive knowledge of the anatomy and an understanding of the surgery and the task at hand are needed. These prerequisites are provided in residency training but not in medical school.

Differences between the groups may not lie simply in mean performance but also and particularly in performance variability. Prior training does not simply improve performance, it also improves consistency. The ES3 is able to provide a training curriculum at the end of which all subjects are able to perform well and no extreme outlier is observed, thus narrowing down the variability of their performance. It is remarkable that the level and consistency of performance on virtual surgical tasks become equivalent between inexperienced medical students and senior otolaryngologists, as emphasized in Figures 2, 3, 4, 5, and 6.

Establishing benchmark criteria by expert surgeons is an important step in being able to measure the ability of the ES3 to train and improve the surgical skills of otolaryngology residents. Our data clearly show that, at least for the novice mode, students and residents initially start out with disparate performances compared with expert surgeons, but they quickly gain the skills needed to achieve scores comparable with those of the expert surgeons. This does not imply that the skills gained are sufficient to perform unsupervised ESS but rather that the ES3 efficiently teaches the limited subset of skills it is able to teach. Thus, the simulator can discriminate between the levels of experience of the trainees and can be considered a partially validated representation of ESS. Although it was our belief, throughout the past decade of research on the ES3, that its 3 modes were only the sequential pieces of a training continuum, it is now apparent in the light of the present results that they are rather a varied set of tools that may be used for multiple applications. The novice mode seems very reliable in assessing gross training experience inequalities and therefore could be thought of as a tool to evaluate a subject's past training, for instance at the outset of a residency program to verify the proficieny of a resident in ESS (and, indirectly, the efficacy of the training program) or at the level of board certification to verify that a trained otolaryngologist has met the standards for safe performance of ESS. The intermediate mode probably is archetypical of the original goals of the development of the ES3 as a surgical trainer, because it seems that most of the surgical tasks are learned on this mode and that the environment most closely resembles an actual ESS procedure. Finally, the advanced mode has more potential as a rehearsal and practice mode for trained subjects. It sets a framework to think of future applications of simulation such as continuing medical education, and future developments such as expert-level rehearsal and patient-specific surgical planning, that are currently considered only as preprototypical research projects.

Few simulators have been studied for their effectiveness in training and evaluation. Only a limited number of developers have thoroughly validated their surgical simulators, mostly by comparing experienced surgeons with those with less or no surgical experience to demonstrate construct validity, showing that the simulator is measuring the skill it is designed to measure. Other teams have reported diverse results, including a knee arthroscopy simulator and a laparoscopic surgery simulator demonstrating significant differences between those with different surgical experience (discriminant validity), whereas others failed to discriminate between those with different levels of ability. The Minimally Invasive Surgical Trainer-VR (MIST-VR) (Mentice AB, Gothenburg, Sweden) laparoscopic simulator has been shown to possess construct validity because experienced surgeons perform better than novices. It is also instructionally effective for a basic task because those who had trained on the MIST-VR were shown to be better at a laparoscopic cutting task than those who had not. Although the MIST-VR is currently the most promising device for direct use in surgical training and probably the only one with validity features as significant as and comparable with the ES3, a very recent study points out the fact that it still does not seem reasonable to use it as a skills assessment device.

Our observations suggest that the ES3 may be a viable assessment tool for various levels of skills, particularly if the assessment is restricted to the use of the novice mode, which seems to be the most discriminant
segment of the curriculum. Possible applications may include objective applicant selection for admission into residency training, trained physician certification and credentialing, as well as continued medical education for practicing physicians. Additional fidelity investigations are obviously needed before the use of a surgical simulator as a skills assessment instrument becomes generalized and possibly mandated by surgical authorities.

The correlation gap that seems to exist between the 3 difficulty modes of the ES3 also led us to review the very structure of its curriculum for future studies. The elimination of the novice mode seems a viable option when the objective is to train the subject for specific surgical tasks, for instance in the framework of a predictive validity assessment: to demonstrate the extent to which the knowledge gained on the simulator transfers to actual operative practice, or “VR to OR” protocol. Moreover, we believe eliminating the novice mode will also unravel much more discriminating capabilities in the intermediate mode.

A predictive validity study of the ES3 is currently under way; after careful selection of a homogeneous group of novice residents, half were randomly assigned to simulation training, whereas the other half are being observed as controls. The first OR performance by all subjects of a standardized ESS task is videorecorded, deidentified, and assessed by a blinded panel of experts using a set of objective surgical metrics. The results of this comprehensive VR to OR protocol, expected to be available soon, will illustrate the potential of a simulator to translate VR skills into actual dexterity in live surgery.

The combination of training specifically to a surgical task, with benchmark criteria established by the observation of expert performance, generates the fundamental paradigm of “criterion-based” or “proficiency-based” training, in which the objective is not to repeat a task a certain number of times or during an arbitrarily determined amount of time but rather to practice until an objective level of proficiency previously established by experts is reached. Ongoing studies are currently testing this model on the ES3 for ESS procedures.

In all studies involving surgical simulator validation, one significant limitation is the population sample size. This is also true for the studies conducted on the ES3, for which trainees were recruited in a specialty that has fewer residents compared with other fields. A clear benefit of larger-scale studies would be, for instance, the ability to decipher fine variations of skills within a population of trainees rather than comparing them with other groups of individuals. However, the multi-institutional implementation of our study design allowed us to reach numbers comparable with the mainstream surgical simulation studies. Although most sample populations remain small, they are nevertheless widely accepted today as standard for the particular setting of simulation studies.

Other limitations pertain to simulation itself as an educational concept: most VR simulators are highly sophisticated, customized computer systems integrating various technological components from the imaging, informatics, and mechanical engineering fields. Therefore, their design, development, use, and maintenance often mandate heavy costs and may limit their accessibility. Research efforts leading to demonstrate their ability to provide efficient training will potentially lead to generalized mass production, decrease in costs, and easier acceptance by surgical educators.

**CONCLUSIONS**

The ES3 has established complete and satisfactory construct validity for the measurement of skills pertaining to ESS. It has also been able to provide benchmark performance data gathered from expert surgeons, which will be used as performance criteria with novice trainees. A part of the simulator curriculum also seems viable to assess surgical skills. These are significant additions to support the already increasing acceptance of surgical simulation for resident training.

Further studies are ongoing to assess the ES3’s predictive validity as a training tool prior to the performance of actual ESS, and to determine the soundness of the concept of criterion-based training as a way to effectively alter and modernize currently accepted surgical education principles.

Submitted for Publication: August 22, 2006; final revision received October 31, 2006; accepted December 1, 2006.

**Author Affiliations:** Department of Otorhinolaryngology–Head and Neck Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY (Drs Fried, Sadoughi, Cuellar, and Uribe and Mr Zeltsan); Human Interface Technology Laboratory (Ms Weghorst) and Department of Surgery (Dr Satava), University of Washington Medical Center, Seattle; Department of Surgery, Section of Otolaryngology, Yale University School of Medicine, New Haven, Conn (Drs Sasaki and Ross); and Department of Otolaryngology, New York University Medical Center, New York (Drs Jacobs and Lebowitz).

**Correspondence:** Marvin P. Fried, MD, Department of Otorhinolaryngology–Head and Neck Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, 3400 Bainbridge Ave, Third Floor, Bronx, NY 10467 (mfried@montefiore.org).

**Author Contributions:** Drs Fried, Sadoughi, Uribe, and Jacobs had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. **Study concept and design:** Fried, Sadoughi, Ross, and Satava. **Acquisition of data:** Sadoughi, Zeltsan, Cuellar, Uribe, Sasaki, Jacobs, and Lebowitz. **Analysis and interpretation of data:** Fried, Sadoughi, Weghorst, Zeltsan, and Uribe. **Drafting of the manuscript:** Sadoughi and Cuellar. **Critical revision of the manuscript for important intellectual content:** Fried, Sadoughi, Weghorst, Zeltsan, Uribe, Sasaki, Ross, Jacobs, Lebowitz, and Satava. **Statistical analysis:** Sadoughi. **Obtained funding:** Fried and Satava. **Administrative, technical, and material support:** Fried, Sadoughi, Weghorst, Zeltsan, Cuellar, Uribe, Sasaki, and Satava. **Study supervision:** Fried, Sadoughi, Cuellar, Ross, Jacobs, and Lebowitz.

**Financial Disclosure:** None reported.

**Funding/Support:** This study was funded by research grant No. R18 HS11866-03 from the Agency for Healthcare Research and Quality.
REFERENCES


