Sinus and Skull Base Surgery Simulator – Development and Validation

by

Allan Dave Vescan

A thesis submitted in conformity with the requirements for the degree of Masters of Science

Institute of Medical Science
University of Toronto

© Copyright by Allan Dave Vescan 2015
Sinus and Skull Base Surgery Simulator – Development and Validation

Allan Dave Vescan

Masters of Science

Institute of Medical Science
University of Toronto

2015

Abstract

The acquisition of surgical skill has evolved dramatically over the last century. There has been a paradigm shift away from an apprenticeship model where learning occurs in the clinical setting on live patients, to a process where basic skills are acquired in a surgical skills center on a spectrum of simulators.

The evolution in the rapid prototyping industry has been slowly integrated in to the medical community. Surgical subspecialists have relied on medical modeling to help with surgical planning.

A novel sinus-skull base simulator was developed and validated. Objective structured assessment of technical skill was able to show construct and concurrent validity. Face validity was achieved. However, surgical tool kinematics did not show construct validity.

The validation of a novel sinus/skull base simulator may be a beneficial, cost controlled tool to allow novice surgeons to gain basic skills in endoscopic sinus/skull base surgery.
Acknowledgments

First and foremost I would like to thank Mike Daly and Harley Chan for their guidance and assistance in the surgical skills laboratory and for all their technical input with respect to imaging acquisition and creation of 3D models/simulators. I would also like to thank Marcela Fandino for her time and assistance at facilitating the simulation trial and acquisition of all of the video assessments. In addition, I would like to thank the members of my Program Advisory Committee, Jon Irish, Ian Witterick and Adam Dubrowski for their valued advice, patience and input in helping a clinical surgeon start to think like a scientist.
Contributions

I would like to thank the following for the specific work presented here. First Mike Daly for his assistance with calculation of all the metrics associated with instrument tool tracking and hand motion analysis. Furthermore, Harley Chan and Thomas Looi for their help with creation of the 3D-models of the sinus and skull base and Dr. Laurel Duquette for her guidance and assistance with statistical analysis. In addition, I would like to thank Dr. Jeff Siewerdsen for his input with the initial stage of process validation of the three-dimensional printing. Finally, I would like to acknowledge the experts, Drs. John Lee, Chris Dhepnorarat, Arif Janjua, Kristian Macdonald and Yvonne Chan who spent numerous hours reviewing video and rating the performance of both novice and expert surgeons.
# Table of Contents

Acknowledgments........................................................................................................... iii

Table of Contents ........................................................................................................... v

List of Tables ................................................................................................................ viii

List of Figures ................................................................................................................ x

List of Abbreviations .................................................................................................... xiii

Chapter 1 ......................................................................................................................... 1

1 Introduction .................................................................................................................. 1
  1.1 Simulation ............................................................................................................... 1
    1.1.1 Introduction ...................................................................................................... 1
    1.1.2 Acquisition of Surgical Skills ........................................................................... 2
    1.1.3 Types of Simulators ....................................................................................... 4
    1.1.4 Assessment Tools .......................................................................................... 8
    1.1.5 Rapid Prototyping ......................................................................................... 15
    1.1.6 Medical Application of Rapid Prototyping .................................................... 20
    1.1.7 Limitations of Rapid Prototyping .................................................................. 23

Chapter 2 ......................................................................................................................... 26

2 Objectives and Hypothesis ......................................................................................... 26

Chapter 3 ......................................................................................................................... 28

3 Development of a Novel High-fidelity Sinus/Skull Base Simulator with the aid of Three-Dimensional Printing ........................................................................ 28
  3.1 Introduction .......................................................................................................... 28
  3.2 Materials and Methods ......................................................................................... 30
    3.2.1 Subject selection and CT Image Acquisition .................................................. 30
    3.2.2 Digital Processing .......................................................................................... 31
    3.2.3 Printing and Material Development ............................................................... 33
    3.2.4 Accuracy and Reproducibility of Material Printing ........................................ 35
Chapter 4

4 Validation of a Sinus/Skull Base Surgery Simulator developed with Three-Dimensional Printing

4.1 Introduction

4.2 Materials and Methods

4.2.1 Subject Recruitment

4.2.2 Simulator Assembly

4.2.3 Surgical Tasks

4.2.4 Assessment Tools

4.2.5 Instrument Tracking/Hand Motion Capture & Analysis

4.2.6 Video Capture and Editing

4.2.7 Statistical Analysis

4.3 Results

4.3.1 Demographics

4.3.2 Level of Realism

4.3.3 OSATS Analysis

4.3.4 Inter-rater Reliability

4.3.5 Instrument Tracking Analysis

4.4 Discussion
List of Tables

**Table 3-1:** Summary of example materials tested for simulation of bony and soft tissues in the prototype simulator………………………………………………………………..Pg. 35

**Table 3-2** Summary of Qualitative Assessment of Example materials……………………… Pg. 42

**Table 3-3** Mean depth and standard deviation of ZP-130 and ZP-15 test samples of membrane thickness……………………………………………………………………..Pg. 46

**Table 4-1** Simulator Global Rating Scale……………………………………………………………Pg. 63

**Table 4-2** Simulator Task Specific Scale (SIM-TSS)………………………………………………Pg. 63

**Table 4-3** Cadaver Task Specific Scale(Cad-TSS)……………………………………………… Pg. 64

**Table 4-4** Level of Realism………………………………………………………………………………Pg. 65

**Table 4-5** Demographic information…………………………………………………………………Pg. 71
Table 4-6 Level of Realism.................................................................Pg. 73

Table 4-7 Summary of the results..........................................................Pg. 74

Table 4-8 Inter-rater reliability..............................................................Pg. 77
List of Figures

**Figure 3-1.** Flowchart illustration of the process for creation of the high-fidelity phantom…………………………………………………………………………………………..Pg. 33

**Figure 3-2** Honeycomb 3D printing test materials…………………………………………Pg. 36

**Figure 3-3** (a) Ten ZP 15 and (b) ten ZP 130 spheres…………………………………………Pg. 38

**Figure 3-4** Feature pattern to test minimum feature size…………………………………Pg. 39

**Figure 3-5** Flexible Zp-15 test plate………………………………………………………..Pg. 41

**Figure 3-6** Modulation in 3DP feature size measured as a function of frequency
(separation)…………………………………………………………………………………………..Pg. 44

**Figure 3-7** Illustration of high-fidelity sinus and skull base prototype for surgical simulation…………………………………………………………………………… Pg. 47

**Figure 3-8** Experimental setup for surgical task performance…………………………..Pg. 48
**Figure 4-1** Assembly of Sinus/Skull Base Simulator........................................Pg. 60

**Figure 4-2** Simulator Assembled........................................................................Pg. 61

**Figure 4-3** (a) Experimental setup for phantom surgical studies. (b) Close up view of the IR tools and EM sensors used for tracking.........................................................Pg. 67

**Figure 4-4** Example of both construct validity and concurrent validity on one of 13 items tested..................................................Pg. 75

**Figure 4-5** Post-hoc Analysis................................................................................Pg. 76

**Figure 4-6** Time Analysis......................................................................................Pg. 78

**Figure 4-7** Tool Separation.................................................................................Pg. 79

**Figure 4-8** Path Length.......................................................................................Pg. 80

**Figure 4-9** Jitter/Motion Smoothness.....................................................................Pg. 81
Figure 4-10 Compactness ................................................................. Pg. 82

Figure 4-11 Compactness by Surgical Experience Level and Task ................................ Pg. 83
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGME</td>
<td>Accreditation Council for Graduate Medical Education</td>
</tr>
<tr>
<td>PGY</td>
<td>Post Graduate Year</td>
</tr>
<tr>
<td>TSS</td>
<td>Task Specific Checklist</td>
</tr>
<tr>
<td>GRS</td>
<td>Global Rating Scale</td>
</tr>
<tr>
<td>OHNS</td>
<td>Otolaryngology – Head &amp; Neck Surgery</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>ES3</td>
<td>Endoscopic Sinus Surgery Simulator</td>
</tr>
<tr>
<td>IRR</td>
<td>Inter-rater Reliability</td>
</tr>
<tr>
<td>OSATS</td>
<td>Objective Structural Assessment of Technical Skill</td>
</tr>
<tr>
<td>HMA</td>
<td>Hand Motion Analysis</td>
</tr>
<tr>
<td>ESS</td>
<td>Endoscopic Sinus Surgery</td>
</tr>
<tr>
<td>ICSAD</td>
<td>Imperial College of Surgeons Assessment Device</td>
</tr>
<tr>
<td>ADEPT</td>
<td>Advanced Endoscopic Psychomotor Trainer</td>
</tr>
<tr>
<td>MIST-VR</td>
<td>Minimally Invasive Surgical Trainer Virtual Reality</td>
</tr>
<tr>
<td>RP</td>
<td>Rapid Prototyping</td>
</tr>
<tr>
<td>MRP</td>
<td>Medical Rapid Prototyping</td>
</tr>
<tr>
<td>SLA</td>
<td>Stereolithography</td>
</tr>
<tr>
<td>FDM</td>
<td>Fusion Deposition Modeling</td>
</tr>
<tr>
<td>SLS</td>
<td>Selective Laser Sintering</td>
</tr>
<tr>
<td>LOM</td>
<td>Laminated Object Manufacturing</td>
</tr>
<tr>
<td>CBBT</td>
<td>C-arm Cone-Beam CT</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
<tr>
<td>EBM</td>
<td>Electron Beam Melting</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>3DP</td>
<td>Three- Dimensional Printing</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging Communications in Medicine</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield Units</td>
</tr>
<tr>
<td>SP</td>
<td>Standardized Patients</td>
</tr>
<tr>
<td>SIM-GRS</td>
<td>Simulator Global Rating Scale</td>
</tr>
<tr>
<td>SIM-TSS</td>
<td>Simulator Task Specific Scale</td>
</tr>
<tr>
<td>CAD-GRS</td>
<td>Cadaver Global Rating Scale</td>
</tr>
<tr>
<td>CAD-TSS</td>
<td>Cadaver Task Specific Scale</td>
</tr>
<tr>
<td>GTx</td>
<td>Guided Therapeutics</td>
</tr>
<tr>
<td>IGS</td>
<td>Image Guided Surgery</td>
</tr>
<tr>
<td>LIVE-IGs</td>
<td>Localized Intraoperative Virtual Endoscopy</td>
</tr>
<tr>
<td>EM</td>
<td>Electromagnetic</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-Class Correlation Coefficient</td>
</tr>
<tr>
<td>FLS</td>
<td>Fundamentals of Laparoscopic Surgery</td>
</tr>
<tr>
<td>ITA</td>
<td>Instrument Tracking Analysis</td>
</tr>
</tbody>
</table>
Chapter 1

1 Introduction

1.1 Simulation

1.1.1 Introduction

Sir William Halstead introduced the traditional concept of medical education to North America from Germany in the late 1800’s. “See one, do one, teach one” is referred to as the Halstedian approach to medical education with the surgical trainee gaining progressive levels of responsibility through an apprenticeship approach. The success of this type of training model is dependent on high volume exposure to patient problems. This high volume of exposure to clinical problems remains a cornerstone of surgical education more than a century after it was originally introduced. There are numerous issues with the delivery of this form of education in today’s current health care climate. These include reduced working hours for residents, fiscal mandates for cost-efficiency in the operating room, increased emphasis on reduction of medical error and increasing complexity of surgical intervention. In 2003, the Accreditation Council for Graduate Medical Education (ACGME) mandated that resident working hours be reduced to 80 hour work weeks commencing on July 1, 2003. This led to a new generation of surgical trainees where the high volume, high work hours model of training would not be feasible. Bridges and Diamond have estimated that one surgical resident over the course of 4 years will cost nearly $50,000 US Dollars in just increased operating room utilization. When this is extrapolated to the number of general surgery residents trained in one year it represents $53 million US dollars/year. Compounded on top of the aforementioned stressors is an added
attention to reduction and elimination of medical errors, which further reduces the ability for surgical educators to mentor and teach trainees in an operating room setting. All of these external stressors have led to a new era in surgical training which starts in the surgical skills laboratory. Basic surgical skills are learned and practiced on models and simulators with the goal being increased proficiency in the operating room setting on actual patients.

1.1.2 Acquisition of Surgical Skills

The ability for an individual to acquire new motor skills and expertise with new styles of training that incorporate simulation is based on well-established theory. Fitts and Posner describe three stages to the acquisition of psychomotor skills that are widely accepted throughout the surgical literature. Kopta, considered one of the pioneers in surgical education, based the three theories of technical skill acquisition on the original concept by Fitts and Posner. The three stages of technical skill acquisition are cognitive, integrative and autonomous. During the cognitive stage, the learner conceptualizes the new skill that they are attempting to learn. They will often seek out information about the task through reading, interacting with their mentors and listening to instruction on correct performance of the task at hand. The task they are learning is still non-fluid and the student will break the task down into a series of subtasks. The next stage is the integrative stage; with practice and feedback forming the main components of this stage. The trainee still needs to consciously think about the steps involved however, the efficiency and fluidity improve throughout this phase of learning. The final stage is known as the autonomous stage where the movements become smooth and seamless. The trainee does not need to think
about each step of the process and the technique becomes automatic. This automation of the
cognitive thought process allows the newly created expert to think about other issues involved in
the procedure and further refine performance. For educators and health care providers, the
ability to take a novice trainee from the early stages of learning in the cognitive phase to the
autonomous stage is of paramount importance. This has the potential to decrease complications
in the operating room and improve actual performance.

Seymour and colleagues 10 looked at 16 residents in a general surgery program that were in PGY
(Post Graduate Year) 1-4. The subjects were randomized to training on a virtual reality
simulator or conventional clinical training. They then performed laparoscopic cholecystectomies
on actual patients and were evaluated on performance by expert reviewers. Conventionally-
trained trainees were 9 times more likely to transiently fail to make progress and 5 times more
likely to injure the gallbladder or burn non-target tissue. Mean errors were 6 times less likely to
appear in the simulator based group. In addition, laparoscopic gallbladder dissection times were
29% quicker in the virtual reality trained group.

Another study looked at 20 general surgery residents who received didactic lectures on
interventional catheter-based techniques. They were then randomized to training on the virtual
simulator versus no further training. They all performed 2 mentored catheter–based
interventions in the angiography suite for lower extremity occlusive disease and were evaluated
by expert surgeons with a task specific checklist (TSS) and global rating scale (GRS). The
trainees that received training on the simulator out performed the other group on both the TSS
and GRS on both the first and second catheter based interventions.11

Another framework has been described by Collins et al.12 to conceptualize the progression of a
novice surgical trainee to an expert surgeon. In this model the novice trainee moves through six
phases from modeling to exploration. The first phase is modeling and is one of the most influential, as staff surgical preceptors play a critical role in all learned behaviours. This includes handling of tissue and surgical technique, interaction with colleagues and nursing staff, and dialogue with a patient’s family members. The next step is coaching. This allows the trainee to form the critical infrastructure and foundation from which to build their knowledge and skills. The feedback from the surgical supervisor involves both encouragement and criticism during this phase. The infrastructure that is built is referred to as the scaffold and serves as the next step, whereby the surgical resident has a thorough knowledge of the procedure. The next phase is articulation, at which point the trainee understands all of the steps of a given procedure and has acquired all of the technical skill in order to perform the operation. They are able to safely complete the procedure independently if all of the factors surrounding the operation are straightforward. The final two steps, reflection and exploration, define the transition from novice to expert. The surgeon can now evaluate their own performance, determine inefficiencies and areas for their own improvement. They start to consider alternative techniques and may even move on to develop new procedures and solutions for pre-existing problems at this stage.

1.1.3 Types of Simulators

There are a wide variety of simulation techniques and bench models that are currently in use by medical and surgical training programs. One of the important concepts when describing simulators is to describe their level of fidelity or realism. Fidelity is often described as being either low fidelity in its relationship to normal human anatomy or high-fidelity, whereby the
simulator or model closely replicates the native anatomy. There are three factors that play a key role in determining fidelity: visual representation, tissue texture and surgical task construct.\textsuperscript{13} Visual representation and tissue texture are self-explanatory. Surgical task construct is the concept that the procedure or maneuver that is being taught is replicated in the model. As the fidelity of the model or simulator increases, the cost associated with production of these simulators also increases.\textsuperscript{1,3,4,14} Interestingly, increased fidelity does not necessarily always translate into improved performance after simulation. Matsumoto and colleagues looked at a low fidelity ureteroscopy and stone extraction model that was constructed for $20 dollars and compared it to a high-fidelity virtual simulator that cost $3700. They took 40 final year medical students and randomized them to one of three interventions. They received either didactic training or hands on training in the low fidelity simulator or high-fidelity simulator. They showed that hands-on training out performed the didactic training students, however, there was no difference in performance between the high-fidelity trained group or the lower fidelity trained students in endo-urological skills. Evaluation of performance was done with video evaluation and rating by experts using GRS and TSS. Anastakis et al\textsuperscript{15} reported similar findings where 23 first year residents were randomized to one of three training methods, low fidelity bench model, cadaver (high-fidelity) or didactic text. A series of six basic surgical tasks were simulated and performance was measured on a cadaver with GRS and TSS after training. There was no difference in performance between the high-fidelity and the low fidelity groups; however, they both outperformed the didactic group. This illustrates an important concept, that level of fidelity does not always translate to improved performance and it is often the task that is being simulated that will dictate the level of fidelity that is required.

Some authors have classified types of simulators as being either organic or inorganic.\textsuperscript{7} Organic forms of simulation include animal models, standardized patients and cadaveric models. These
models are excellent for displaying and practicing techniques of hemostasis and tissue dissection. Their limitations include cost, issues with storage, and specialized personnel that are required to house and take care of the animals. In addition, they are single use and there are ethical concerns around the usage of animals for simulation and training. Although animal models are high-fidelity, the anatomy may not be representative of human procedures. Cadaver specimens are the only “true” anatomical simulator. The main issues with cadaver samples include availability, cost, storage and the properties of the tissue. In addition, they are single use and in certain parts of the world, use of cadaveric dissection is viewed as unethical. There are also concerns about the potential for transmittable infectious disease.

Inorganic simulators include bench models, virtual reality simulators and human performance simulators. Bench models are usually low cost, low fidelity types of simulators that focus in on one task or component of an operation as opposed to the whole procedure itself. They are easy to use, portable and often carry minimal risk. Bench models or “part task trainers” as they have been called are best used with novice trainees. One example of a “part task trainer” would be the Quinsy Trainer which focuses on drainage of a peri-tonsillar abscess. The construct is that of a mannequin head with a water balloon situated in the oropharynx. The trainee is then instructed to drain the collection/water balloon with a needle aspiration. The model is re-usable, inexpensive and transferable.

Virtual reality devices use microprocessors to create an environment where both tissue and surgical instrumentation are computer generated. They have the ability to provide haptic feedback to the trainee. The main advantages of the virtual reality system are the ability to provide detailed assessment of surgical performance with computer generated metrics, in addition they are re-usable and easy to set up. The major limitations to the use of virtual reality
Simulators are cost and variety. Most virtual reality simulators that have been developed have been for laparoscopic use, and therefore, smaller surgical sub-specialties with unique procedures will not be represented by current VR simulators. Recently there have been a number of virtual reality simulators that have been developed for use in the field of Otolaryngology-Head and Neck Surgery (OHNS). These include VOXEL-MAN TempoSurg (VOXEL-MAN Group, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany) and the VOXEL-MAN SinuSurg. The TempoSurg simulator is directed toward surgery of the middle ear and temporal bone and is based on high resolution computed tomography (CT) scans. Three dimensional virtual images are synthesized from the patient images. The SinuSurg simulator is directed toward surgical procedures of the paranasal sinuses and virtual images are created in much the same way as the TempoSurg. Other virtual reality simulators directed toward ear and temporal bone surgery simulation include Mediseus SDS3000 temporal bone simulator (Mediseus, Australia) and the Visible Ear Simulator by Peter Trier Mikkelsen (Alexandra Institute, Aarhus, Denmark) and Mads Solvsten Sorensen (ENT Department, Rigshospitalet, Copenhagen, Denmark). The most extensively studied and evaluated simulator within OHNS is the ES3 Endoscopic Sinus Surgery Simulator (Lockheed Martin, Akron, Ohio). This simulator provides a very realistic two dimensional simulation of sinus surgery on a video display.

The final type of inorganic simulation is the human performance simulators also referred to as integrated simulators. This type of simulation is most commonly used in critical care medicine and anesthesiology simulation. They are excellent for crisis management and team training. Often a mannequin is linked with a complex set of computerized controls that can manipulate physiologic parameters. This would include both physical (heart rate, respiratory rate) and electrical outputs (monitor outputs). An example of an integrated simulator would be the SimMan (Laerdel). This type of simulator is used to recreate clinical scenarios, whereby one
could train both OHNS surgeons and anesthetists to work together during difficult “shared airway” scenarios.

1.1.4 Assessment Tools

1.1.4.1 Validation & Reliability

Simulators have been developed with the goal of being used as both training devices in addition to assessment tools. Prior to the adoption of any new assessment tool, two primary criteria need to be met. These are validity and reliability. Validity is defined as the “property of being true, correct and in conformity with reality”. One of the key principles of any test or assessment tool is that it in fact measures what it was intended to measure. \(^8,25\) In the context of validity this takes on a number of different components. These components are used to establish benchmarks of validity for a new test or assessment tool. They include face validity, content validity, construct validity and criterion validity. \(^9,26\) Criterion validity can be further broken down into predictive validity and concurrent validity. Face validity or realism simply is the concept of whether the test or assessment tool resembles the real life scenario. Content validity refers to the extent the measurement reflects the real activity that is being tested. For instance testing a surgical resident on performing an ethmoidectomy on a cadaveric head has higher content validity than a written examination on the steps of a surgical ethmoidectomy. Construct validity relates to the ability of an assessment tool to differentiate based on levels of experience. For example, to show construct validity of a new simulator’s ability to assess technical performance, expert surgeons should score higher on the metrics used for evaluation compared with novice surgeons. \(^10,20\) The final component of validation is criterion validity which is made up of both predictive and
concurrent validity. Predictive validity is the most valuable form of validation in surgical simulation. An example would be the performance scores of residents performing a specific skill in the surgical skills lab correlating significantly with their performance in the operating room 4 months later. Concurrent validity is the measure of performance on a new test as it relates to an established test on the same construct. An example of concurrent validity would be a surgical trainee’s test score on their ability to perform a laparoscopic cholecystectomy on an animal model, correlating significantly with their performance on a new virtual reality simulator. Van Nortwick et al looked at 83 publications that described their methodology for establishing validity in surgical simulation studies. They found that 60% of studies targeted construct validity, 24% of studies examined concurrent validity and only 5% of studies actually reported on predictive validity results. Overall, 82% of studies were conducted at a single institution with a mean number of 37 participants. In terms of metrics, 34% were simulator generated metrics, 33% were from human evaluations of performance, 6% of metrics used motion analysis and 24% used a combination of the above.

Reliability is the second critical component of any assessment tool and is defined as the ability of a test to generate similar results when applied at two different points in time, with minimal associated errors in measurement. Reliability is often measured as inter-rater reliability (IRR). A widely used rule of thumb is that a reliability test should not be used if the IRR value is <0.7 and it should not be used to form important decisions about an individual unless the IRR value is >0.9.

1.1.4.2 Objective Surgical Skills Assessment
The historical pattern of surgical skill assessment was riddled with problems, including lack of reliability and validity. It included techniques such as procedure lists with logs, which clearly lack validity, as simply tracking the number of cases that a surgical trainee performs is no measure of how well they perform the procedure. Direct observation without criteria has also been reported, however, due to the subjective nature of the evaluation by an individual surgeon the results are notoriously inaccurate and lack inter-observer reliability. For these reasons, there has been development of reliable and valid tools and systems to measure technical skill on both low and high-fidelity simulators. These include Objective Structured Assessment of Technical Skill (OSATS), hand motion analysis (HMA), virtual reality and haptic systems and final product analysis.

OSATS, as originally designed, is based on a 6 station model where trainees perform tasks on a bench model or live animal and are evaluated by expert evaluators with both a global rating scale and task specific check list. The global scale consists of 7 general components of operative skill that are evaluated on a 5 point Likert scale. The task specific checklist includes components that are relevant to the procedure at hand and are marked as yes/no with respect to completion of the task. The drawbacks of the OSATS format of evaluations is that it requires multiple faculty to provide the review and evaluation at the time of testing or if the procedure is an endoscopic procedure where there is digital capture of information, many hours of watching and rating video is required. OSATS has been further modified away from the original concept of a six station bench model or animal model. Lin and colleagues have developed an Endoscopic Sinus Surgery (ESS) OSATS tool that has been modified and validated from the previous work done by Reznick and colleagues at the University of Toronto. It consists of an 8 item TSS which has a 5 point Likert scale where the behavioural anchor that was determined to be an acceptable “pass” was placed in the middle of the Likert scale at three out of five. The rationale for this was to
allow for residents who are below the minimally acceptable pass level to have an opportunity to improve in their skills. In addition, a 10 item GRS also modified from Reznick et al.\textsuperscript{2,12} was used to capture global performance in 9 different domains and an overall score. Laeeq et al.\textsuperscript{3,4,34} went on to show construct validity using the ESS GRS and TSS in video based cadaveric evaluation of sinus surgery. They were able to show that the tool was able to differentiate performance of senior level trainees (PGY 4,5) from junior level residents (PGY 1-3).

Another form of objective evaluation of surgical skill involves hand motion analysis. This type of analysis is adapted from the field of kinesiology and the study of human movement. The most widely reported form of hand motion analysis involves the use of the Imperial College of Surgeons Assessment Device (ICSAD), which uses a commercially available electromagnetic tracking system (Isotrak II, Polhemus Inc, Vermont). The system involves an electromagnetic field generator and two sensors that are placed on the dorsum of the surgeon’s hands. The positional data and co-ordinates are then captured and converted by Bespoke software (Bespoke Software Inc, Clifton NY) into three output measures: total time, number of movements and path length.\textsuperscript{5,35-38} Datta and colleagues\textsuperscript{6,20-23,38} were able to show that hand motion analysis using the Isotrak II system was able to differentiate between levels of experience on a simple vascular bench model that involved suturing a vein graft to an artery. Specifically, number of movements made, total time taken and global rating scores were all predictive of experience level. They also reported a significant score between number of movements made and global scores, however, the task specific checklist had no predictive value of experience level in this model. Smith et al.\textsuperscript{7,24,39} demonstrated that hand motion analysis was also able to discriminate experience level in a porcine model of laparoscopic cholecystectomy. Fifteen surgeons were divided into one of 3 experience groups: less than 10 previous laparoscopic cholecystectomies, 10-100 or >100 procedures. There was a significant difference in performance level as measured by hand
motion analysis between the three groups. The novice surgeons had the largest improvement in performance between the three groups, as one would expect, as the task was repeated 3 times by each subject.

One of the drawbacks of the electromagnetic system is the interference that can result from metallic surgical instruments being placed near the generator or the sensor. In addition, some of the instruments themselves (cautery, drill) have the potential for interference and thus will compromise the ability to acquire data or the integrity of the data which is attained.\textsuperscript{8,40}

The previously mentioned issues with electromagnetic tracking systems has led to the development of optical tracking systems that are reliant on infrared transmission of information. One such tool that has been designed and reported on for measuring dexterity is the “Advanced Endoscopic Psychomotor Trainer” (ADEPT). The original intention of its design was to provide a tool to select residents for endoscopic/laparoscopic surgery. The system involves a static dome that encloses a defined workspace with two laparoscopic instruments. The system has the ability to measure total time taken, angular path length, successful task completion and instrument error scores.\textsuperscript{9,41-43} The ADEPT system has shown that experienced surgeons perform with significantly lower error rates than novice surgeons.\textsuperscript{10,42} A limitation of this system is that it is not transferable to other procedures aside from laparoscopic work. More recently Sonnadara and colleagues have reported on an optical tracking system which is low cost and based on the infrared tracking cameras from Wii remote controls from the gaming industry (Nintendo Inc, Kyoto, Japan). They took 10 experienced surgeons and 10 novice surgeons and had them perform a basic suturing task on a bench model. Data was collected with both a conventional Polhemus Patriot EM system (Polhemus, Vermont) and the new novel optical tracking system. Both EM and optical tracking sensors were placed on the dorsum of the hands. They were able
to show that the system could differentiate novices from experts on both path length and number of movements and thus show construct validity. In addition, they showed that the optical tracking system results correlated with the electromagnetic system results and video global rating scores.

Virtual reality has been defined as a collection of technologies that allow people to interact efficiently with 3D computerized databases in real time by using their natural senses and skills. Surgical virtual reality systems will have a construct such as a laparoscopic frame or a model of a head that will allow the user to interact with the simulator. These virtual reality systems are computer based models that have the ability to generate output metrics. In an international workshop on Metrics for Objective Assessment of Surgical Skill, a group of experts reviewed all assessment methods and metrics and suggested a number of components that should be included in the output metrics that are being used to assess technical skill during simulation. The components include, economy of motion, absence of motion, path length, time to completion, force measurements, error calculations, response latency (time to recover from error) and state analysis (still vs. moving). The Minimally Invasive Surgical Trainer, Virtual Reality (MIST-VR, Mentice Medical Simulation, Gothenburg, Sweden) system is an example of one of the earliest virtual reality laparoscopic simulators that was intended for use as a low-fidelity task trainer. The collaboration between surgeons and psychologists allowed for task analysis of laparoscopic cholecystectomy. This then allowed for the creation of a list of skills that were required to perform the procedure successfully. The list of skills were then translated into a series of three dimensional images of shapes which the user then learns to manipulate. The MIST-VR has six virtual tasks that increase in complexity. The output metrics for the MIST-VR trainer include: task completion time, errors, economy of motion of the left instrument, economy of motion of the right instrument and economy of diathermy instrument.
during the final two tasks. Previous studies have shown that the MIST-VR has construct validity and can differentiate between novice and expert surgeons. \(^7,^{47}\)

Within the field of otolaryngology-head and neck surgery, the virtual reality simulator that has been studied most extensively is the ES3 Endoscopic Sinus Surgery Simulator. It consists of CT derived 3D paranasal sinus models and interactions with a virtual endoscopic instrument with haptic feedback in addition to a replica of an endoscope and a rubber headed mannequin. \(^16,^{20}\)

There are 8 different surgical tasks and three different levels of difficulty that can be set. The novice mode contains 3 tasks: navigating in a virtual space through 4 sets of hoops, injection of local anesthesia into 5 different targets and dissection of 12 3D structures. In intermediate and advanced modes there are 8 tasks that include: scope navigation through loops, injection of local anesthetic in 3 subsites (middle turbinate, lateral nasal wall and nasal passage), middle turbinate medialization, uncinectomy, polypectomy, anterior and posterior ethmoidectomy, agger nasi cell dissection and maxillary antrostomy. In contrast to the output metrics that are described in the MIST-VR system, the ES3 records overall and task specific scores for each task based on time to completion and accuracy relative to optimal performance. If the subject takes excessive time to complete the task or surgical warning zones have been inadvertently disrupted the total and task specific scores will be penalized. \(^17,^{21}\)

Fried and colleagues have shown that the ES3 has construct validity and can in fact distinguish between levels of experience during certain modes of operation.

Final product analysis is another type of objective measure of technical proficiency of task performance. As the name implies it involves taking the final product that the trainee/subject creates and having expert reviewers rate the quality of the construct. Szalay et al \(^7,^{48}\) looked at 20 surgical residents who took part in a 6 station bench-model examination. They had 2 expert
reviewers per station, for a total of 12 expert reviewers who reviewed the final product from the examinees at each station. The final product was analyzed with a four item five point anchored rating scale which assessed completeness, esthetics, function and overall quality. They found that the method showed construct validity as it was able to differentiate between level of experience and they also found a correlation between OSATS and the final product assessment methods. In another example of final product analysis Datta and colleagues evaluated 6 surgical trainees in a vascular anastamosis model. They compared final product analysis with electromagnetic motion analysis results. In this study final product analysis consisted of a 4 item global rating: suture spacing, suture eversion, quality of the anastomosis heal and quality of the anastomosis toe. They found a significant correlation between manual dexterity as measured by electromagnetic motion analysis and final product analysis as it pertained to two outcome measures: diamater of anastamosis and leak rates. They concluded that it may be possible to predict surgical outcomes based on assessment of technical skill. 18,35

1.1.5 Rapid Prototyping

Rapid prototyping (RP) is the process by which a computer generated 3D model is converted to a physical 3D model. It has been used by many industries to develop high-fidelity models; this includes the automobile, toy and computer industries to name a few. The goal is to develop a model that can be analyzed and modified prior to initiating production. 19,49 The main advantage of rapid prototyping is to take a process that usually would be extremely laborious such as mould making and casting and turn it into a process that would only take hours to produce a desired object. 20-23,50 Medical rapid prototyping (MRP) is defined as the “manufacture of dimensionally
accurate physical models of human anatomy derived from medical image data using a variety of rapid prototyping technologies". The source of the medical image data has been predominantly derived from computed tomography, however, there have also been reports of magnetic resonance and ultrasound based prototypes. The primary tissue type that is produced with the rapid prototyping is a hard model that represents bone, however, models based on soft tissue which includes brain, skin and nasal passageway have also successfully been modeled. The first descriptions of MRP date back to 1990 with anatomic 3D model constructs. Over the last 20 years there have been dramatic advancements in imaging technology with developments in 3D image processing software, computer hardware and prototyping devices that has led to more widespread adoption of MRP. In a European multicenter study looking at the MRP application of stereolithography, 172 respondents listed the following as applications for RP in their institutions: templates for surgical resection, to enhance pre-operative simulation, to aid in the production of a surgical implant, to improve surgical planning, to enhance diagnostic quality, to help with navigation during surgery and finally to assist in counseling of patients preoperatively. The respondents replied that the most common diagnoses for the stereolithography application included: neoplasms, congenital disease, trauma, maxillo-facial anomalies and others. Interestingly, MRP for the purpose of surgical simulation and education was not one of the responses by the individuals in this study.

There are five major types of rapid prototyping technologies: stereolithography (SLA), fusion deposition modeling (FDM), selective laser sintering (SLS), laminated object manufacturing (LOM) and finally 3D printing.
1.1.5.1 Stereolithography

Stereolithography (SLA) was developed in 1986 by 3D Systems (3D Systems, Rock Hill, SC, USA). The SLA RP process involves three primary components: a bath of photosensitive resin, a platform for the model construct to be built on and a UV laser to cure the resin. The layers of the 3D model are built up sequentially from the bottom up through a sequence of curing with the UV laser that is being directed by a mirror into the resin and lowering into the photosensitive bath repeatedly. Once the model is constructed it is then stored in a UV cabinet for a pre-defined time period to further augment the curing process. There are often support structures that act as a scaffold to stabilize parts of the 3D construct that might have an overhang. These must be removed by hand in an often laborious and time consuming process. The material used in SLA is a photopolymer. There are a number of commercially available photopolymers and all of them are a type of acrylate. \(^{11,27,49}\)

1.1.5.2 Fused Deposition Modeling

FDM is similar to SLA from the perspective of a layered construction, however, instead of curing a photopolymer with a UV laser, the process involves extrusion of a thermoplastic through a fine nozzle. Once the layer is sprayed the nozzle will rise and the next layer will be sprayed on. As with SLA there is a requirement for supporting scaffold for overhanging components of the model, due to the fact it often takes time for the thermoplastic to harden. The scaffold will be sprayed on by a second jet and often will be done in a second color. This also requires manual removal that is time intensive post production. A commonly used material
during the FDM process is acrylonitrile butadiene styrene, however, other materials can be used including a proprietary nylon and investment casting wax. There is no post-production curing required.12,28,49

1.1.5.3 Selective Laser Sintering

SLS is based on a carbon dioxide laser that sinters successive layers of a powder instead of liquid. The process of sintering involves forming a solid mass of material by heat and/or pressure without melting it to the point of liquefaction. The object that is being created will sit on a platform that will support the construct and is surrounded by the powder that will be sintered. After one layer has been constructed a roller will roll the next layer of powder over the layer that has just been processed, and ultimately a 3D prototype will be created. One of the advantages of the SLS technique is that it does not require any supporting structures to be created as the model is in a supporting powder during processing. Therefore, no post processing manual labor is required, significantly reducing production costs. Materials that can be used include thermoplastic, metal, ceramic or glass powders that can fused. In addition, polycarbonate and polyvinyl chloride are options.13,29,49.

1.1.5.4 Laminated Object Manufacturing

The premise behind LOM is that there is a supply roll of material and an uptake roll of material. As the supply roll unravels it passes over a platform that will support the construct being created. While the roll material is overtop of the platform, a carbon dioxide laser will follow the contour
of the predesigned construct and cut out the shape that is required. Each successive layer that is created will be fused to the previous layer either by welding or an adhesive which can be applied. There is a large array of sheet materials that can be used for production including: papers, metals, plastics, fibres, synthetic material, glass or composites. Due to the fact that no curing is required and the only processing involves cutting the materials, LOM is a comparatively quicker process. The sheet materials are inexpensive which makes LOM a favourable technique for large volume modeling. However, with respect to MRP, LOM has significant issues as it is not possible to remove the solid waste from the hollow cavities and spaces that are often present in human anatomy.

1.1.5.5 Three Dimensional Printing (3DP)

Three dimensional printing was developed at the Massachusetts Institute of Technology in the early 1990’s. The process of 3DP begins with 2-D deconstruction of a 3D model by a computer software program. A layer of powder is then laid down on top of a platform which is supported by a piston within a cylindrical casing. The next step involves an inkjet printer head that projects droplets of binder fluid down on to the powder at spots that correspond to areas where solidification is desired. The process repeats itself layer by layer until a 3D model is created. The powder that is used for printing is either a plaster based or starch based powder. It surrounds the printing object and acts as a support throughout the printing process, it needs to be removed following printing manually. Some of the advantages of the inkjet based 3D printing include low cost, high speed and multimaterial capability.
1.1.6 Medical Application of Rapid Prototyping

While rapid prototyping has been well established in the manufacturing industries, it has been over the past ten to fifteen years where it has gained some traction within the field of medicine. It’s usage within medicine can fit into one of three categories: individual patient care, research and medical education and training.

1.1.6.1 Individual Patient Care

Rapid prototyping has been used in numerous fields of surgery including maxillofacial, craniofacial, orthopedic, neuro and cardiovascular. Its main application revolves around pre-operative planning and to aid in the understanding of complex anatomic problems. Its use in maxillofacial surgery and craniofacial surgery has been well documented and the advantages well described, including planning of orthognathic procedures, assistance with tumor resection and implant placement and shaping. The primary uses in orthopedic surgery have been for preoperative surgical planning, including spinal surgery and planning of screw placement. Other applications include complete custom implants and novel new uses such as bioscaffolds for bone and tissue engineering. In the field of neurosurgery it has been used to plan and aid in the understanding of complex anatomy associated with aneurysm surgery and in certain congenital malformations of the ventricular system.

Other uses that have been described for individual patient care include implant and tissue design. The implants and tissues can be divided into use for bone or soft tissue reconstruction. One of
the more remarkable examples of this is auricular reconstruction. Patients with anotia that have not developed an external ear can have the contralateral ear prototyped and through computerized mirroring, an ear on the side that corresponds with the anotia can be created. This 3D prototype then can be used to serve as a model to create the actual prosthesis in the compound that most resembles native soft tissue. In the area of bony implants the primary sites of prototyping application are in mandibular, maxillary, hip, hemi-knee joint reconstruction and dental restoration.

1.1.6.2 Medical Education and Training

An extensive knowledge of human anatomy is fundamental to being a skilled surgeon. Specifically, surgical procedures rely on a comprehensive understanding of 3D anatomic relationships. The current form of education for most medical students involves didactic lectures and cadaveric dissections in the early years of medical training. This foundation is then built upon through an apprenticeship model during early clinical years in surgical training programs. The opportunity to augment this traditional education model can enhance student knowledge and skill. Rapid prototyping allows the creation of models that help students understand complex 3D anatomy. One example of this is a temporal bone model that was produced by Suzuki and colleagues. This temporal bone model was created with SLS and allowed for simulation of a number of surgical maneuvers that are performed with ear surgery. The ability to recreate 3D anatomy that is not displayed on a monitor but can actually be grasped and manipulated has the potential to improve a surgeon’s abilities and results by allowing them to pre-select the best treatment decision before an operation or procedure. Another potential advantage to rapid
prototyping is that young surgeons can have a simulated environment with high intensity training where errors and refinement of skill can take place without any risk to patients.  

1.1.6.3 Medical Research

In the arena of medical research, rapid prototyping has allowed for the creation of 3D models that can be used to recreate systems to help understand physiologic processes that are still being discovered. One example of this is the use of RP to create models to investigate hemodynamics and optical flow measurements. de Zelicourt et al used SLA techniques to create transparent models to investigate patient-specific total cavopulmonary connection morphologies. The use of SLA techniques significantly reduced the labour involved in creating models required for investigation of digital particle image velocimetry, laser Doppler velocimetry and flow visualization. Another example of more clinically relevant research with rapid prototyping is a study done by Prisman et al where ten patients who were to undergo mandibulectomy for oncologic resection had their mandibles prototyped preoperatively. Using FDM technique for prototyping the reconstruction plates required for the mandible were contoured both pre-operatively and intra-operatively. Two expert surgeons who were blinded to the guide for plate contouring (patient mandible vs prototype mandible) reviewed and rated the plates based on a three question evaluation tool with a five point Likert scale. There was no significant difference in the plates produced by either technique. However, the pre-operative plate was chosen for use the majority of the time and the time saved intraoperatively was significant as the process of contouring a plate during surgery can take up to fifteen minutes. In another study by Daly et al a RP model from FDM techniques was used to establish a system that would allow for overlay
of an augmented reality video output onto traditional endoscopic video images. This novel surgical dashboard has further been evaluated for its ability to help surgeons, both novice and expert, in their ability to identify complex endoscopic skull base anatomy and help with surgical precision.\textsuperscript{82,83} The ability to recreate human anatomical models with RP technology presents the opportunity for many research related topics ranging from creation of physiologic models to simulation and new technology evaluation.

1.1.7 Limitations of Rapid Prototyping

There are a number of limitations and common pitfalls that have been described for RP as it pertains to its application for medicine. They can be divided into global issues or technical pitfalls. With respect to global issues one of the limiting factors is print size. Currently RP techniques are not able to print whole body models. This necessitates the fabrication of smaller parts that then can be assembled and post processed in order to achieve desired size. The largest size that can be produced is by SLA techniques at 20 x24” while FDM is limited to 12 x 12”. The other techniques of 3DP, LOM and SLS fall somewhere in between these ranges.\textsuperscript{49} Other global issues with respect to RP technology include time and cost for production. Given the labor involved with printing and processing some of the models, production time can vary upwards of 1-2 weeks. In the setting of operative planning this would be of value only in planned elective surgery where there is sufficient lead time to take all the necessary steps to create a model for planning or education. However, as one would expect in the setting of an emergency procedure this would not be a feasible solution.\textsuperscript{84} Cost is also potentially a prohibitive factor, with some models including labor estimated at $700-$1400 US dollars.\textsuperscript{84}
Material safety and selection can also be considered a factor in the efficacy of RP. Only a few RP materials are considered safe for transport into the operating theater and in a previous publication Frame and Huntley commented that none at the time were safe for use in the human body. The final global limitation that has been described revolves around “ease of use”. RP machines require an established set of technical skill and experience in order to produce high quality medical models. In addition to the actual process of printing and post processing of the models, a significant amount of knowledge is required for the software programming and data preparation. These factors have a significant impact on medical facilities that are considering adoption of RP technology as an investment in training and staffing in addition to capital investments will need to be strongly considered.

The second component to rapid prototyping pitfalls in medicine involves the production process. Winder and Bibb reported on their six year experience with over 350 different medical models and reported the following as common concerns in production of models primarily using SLA and FDM techniques. They noted that a number of factors can create artifact that leads to distortion on the medical models and includes: CT data import errors, CT gantry tilt distortion, metal and motion artifact, surface roughness due to support structure removal or surface modeling and image data thresholding issues. Image data thresholding problems have the potential to omit large areas of anatomical interest for example in the region of the bone of the medial orbit. For this reason and the previously mentioned ones the authors recommend a multidisciplinary approach to medical model construction where the technician, engineers and clinicians work closely together to create accurate models.
Chapter 2

Objectives and Hypothesis

A validated sinus / skull base simulator that allows acquisition of basic endoscopic surgical skill, outside of real time clinical encounters with patients, would greatly enhance the ability to train novice otolaryngology-head and neck surgery residents. The objective is to develop a realistic high-fidelity simulator based on true anatomic modeling from patient computed tomography imaging. This simulator would allow for replication of surgical maneuvers on an inanimate model that could be reproduced and evaluated. One of the main advantages of producing a simulator based on rapid prototyping technology, specifically 3D printing, is significant cost reduction over existing virtual reality simulators for endoscopic sinus surgery. \(^{19,86}\)

A secondary objective of this study is to use the newly developed sinus / skull base simulator to determine whether novel surgical instrument kinematic metrics that we developed are a valid and reliable tool for assessing level of expertise. The simulator will help evaluate the validity and reliability of these new objective measures of surgical performance, in addition to, other metrics that may be introduced in the future.

This thesis presents two studies based on the development of the sinus/skull base simulator for training and assessment of surgical skill during endoscopic sinus and skull base surgery.

The first study (Chapter 3) describes the development and characterization of the simulator. It looks at the accuracy of the three dimensional printing process with a focus on material selection and realistic surgical feel and imaging properties.
The hypothesis is that a realistic sinus / skull base simulator can be developed that accurately resembles normal human anatomy and it can be used to simulate technical surgical steps that would be encountered during sinus and skull base surgery.

The second study (Chapter 4) will evaluate surgical performance on the new simulator, comparing different levels of expertise and performance with previously validated, established simulation tools (cadaver). Objective metrics for evaluation of surgical performance will include blinded expert video analysis of the tasks completed on the simulator (OSATS). A secondary goal will be to determine the validity and reliability of novel surgical tool kinematics.

The hypothesis of this study is that the sinus and skull base simulator will be able to differentiate novice from expert surgeons (construct validity) using established methods of objective evaluation and with new novel surgical tool kinematics. In addition, performance on the simulator will be equal to performance on established forms of simulation (cadaver dissection, concurrent validity).

After the presentation of the two studies a general discussion will follow discussing the merits of process validation of a sinus/skull base simulator developed from 3D printing. In addition, the objective validation of the simulator using video analysis and surgical instrument kinematics will be explored.

Finally, a review of surgical simulation, future directions and the importance to surgical education and evaluation will be discussed. This will include ways to improve on current approaches to training surgical residents in endoscopic sinus surgery and future applications of three dimensional printing in otolaryngology-head and neck surgery training programs.
Chapter 3

3 Development of a Novel High-fidelity Sinus/Skull Base Simulator with the aid of Three-Dimensional Printing.

This article was published in Medical Imaging 2009: Visualization, Image-Guided Procedures and Modeling.

3.1 Introduction

Skull base surgery has evolved over recent years from traditional open approaches to minimally invasive endoscopic approaches. With the evolution has come increased risk and potential for patient morbidity even for the most experienced skull base surgeons. Critical anatomical structures such as the optic nerves, carotid arteries, and pituitary gland are found within millimeters of each other in the central skull base. This evolution of surgical technique has necessitated an evolution in technical support during surgery, including image guidance and intraoperative image acquisition. A key example is the recent development of C-arm cone-beam CT (CBCT) for guidance of high-precision head and neck surgery. Surgical expertise and experience has traditionally been acquired through extended training programs based either on an apprenticeship type training model and/or on cadaveric dissection. Recently, there has been a trend toward maximizing efficiency of training programs through introduction of simulation-based training. Simulation can occur through a number of different pathways including virtual reality (VR) simulation using computer modeling to simulate surgical procedures, in addition to both low-fidelity and high-fidelity based surgical models. An advantage of high-
fidelity simulators is increased realism and the haptic feedback during task completion. In addition, the simulator is a replica of the surgical site encountered during actual patient encounters thus more closely recreating the surgical experience.

The main challenge with development of high-fidelity surgical simulation is the inherent complexity of attempting to model human anatomy. This is where recent advances in rapid prototyping technology, such as 3D printing, have begun to change the surgical simulation landscape. 3D printing has evolved to allow the creation of timely, low cost prototypes that can be designed from patient CT data. Rapid prototyping of such models involves a variety of potential techniques, including: selective laser sintering (SLS), fused deposition modeling (FDM), stereolithography (SLA), laminated object manufacturing (LOM), electron beam melting (EBM,) and three-dimensional printing (3DP). Many of the aforementioned techniques require significant time, skilled labor, and cost. The main advantage of 3DP is the ability to produce quick, inexpensive prototypes in a relatively simple manner. 95,96

The purpose and motivation of the work reported below is to develop a high-fidelity simulator that allows for objective evaluation of specific surgical tasks associated with endoscopic sinus and skull base surgery. Specific tasks include landmark identification, middle turbinate resection, quilting suture of the nasal septum, bi-manual drilling of the sphenoid sinus and identification of the sella turcica – each conducted in proximity to simulated normal critical anatomy, such as the optic nerves, carotid artery, and central nervous system.

This paper reports on the development of a high-fidelity three-dimensional prototype for paranasal sinus and skull base surgery simulation.
The technical aspects involved in prototype design and fabrication are discussed, including acquisition of CT images, digital processing, and conversion of image data to printer-ready data format.

As 3DP can employ a number of different materials, each with its own physical properties and imaging characteristics, the development and selection of 3DP materials accurately simulating bony and soft tissues is discussed.

Process validation was conducted to determine the reproducibility, accuracy, and spatial resolution of the 3DP process – each critical to understanding the limitations of 3DP for the development of sophisticated anatomical models.

3.2 Materials and Methods

3.2.1 Subject selection and CT Image Acquisition

Figure 3-1 illustrates the overall process of phantom creation from CT scanning to final assembly. The initial step involves acquisition of high resolution CT scans of either a patient or cadaveric head. For the initial prototype, we chose to use a cadaveric specimen with no known abnormal disease or anatomical variation. Cadaveric specimens inherently exhibit well pneumatized paranasal sinuses and very little evidence of inflammatory changes. The cadaveric head was scanned on a GE Lightspeed 16-slice CT scanner (GE Healthcare, Milwaukee, WI) with isotropic (1 mm) voxel size. For actual patient scans, it is important to consider that inflammation or swelling of the lining of the nasal passages and sinuses would create narrowing of the nasal corridor and thus inhibit the ability to insert instrumentation into the simulator.
3.2.2 Digital Processing

Once the initial CT scan is attained, the digital imaging and communications in medicine (DICOM) data are imported to 3D visualization software (Mimics v12.0, Materialise, Ann Arbor MI). The initial prototype involved definition of two distinct types of structures – “bony” and “soft” tissue – each with distinct material considerations. Structures of interest were segmented...
using a combination of automatic and semi-automatic intensity thresholding, region growing, and slice-by-slice manual contouring in orthogonal planes (axial, sagittal, and coronal). The components of the sinuses and skull base defined as “bony” structures included: the frontal, ethmoid, maxillary, and sphenoid sinuses, the nasal bones, maxilla, orbital walls, palate and skull base. Components defined as “soft tissue” structures included: the septum, middle turbinate, and inferior turbinates. As discussed below, the “bony” and “soft tissue” structures require material selection and processing appropriate to the simulated tissue characteristics. The bony structures are intended to be of density (Hounsfield units (HU)) equivalent to bone and present realistic fracture characteristics and drilling purchase. The soft-tissue structures are intended to be pliable and non-brittle, allowing realistic excision using common surgical instruments, such as straight wildes forceps and through-cuts (Medtronic Xomed, Jacksonville, FL). As illustrated in Fig. 3-1, these segmentations formed the basic scaffold of the high-fidelity sinus simulator.

Following segmentation, the 3D structures are refined (smoothed and noise-reduced), and support units may be digitally added. For example, the prototype incorporated ball joints added to the turbinates that could be snapped into holes defined in the surrounding bony material. Similarly, a slot was created along the top and bottom edge of the septum that allows the septum to be slid into the phantom in final assembly. Following such digital processing of the segmented structures, the data are converted to Standard Tessellation Language STL file format (triangular mesh) common to CAD and rapid prototyping technology.
3.2.3 Printing and Material Development

In the development of the initial prototype phantom, we evaluated material selection and post-processing techniques, with 12 materials summarized in Table 1.

<table>
<thead>
<tr>
<th>Module #</th>
<th>Material</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Polycarbonate</td>
<td>None</td>
</tr>
<tr>
<td>#2</td>
<td>Z-Bond</td>
<td>1 CA101 dip</td>
</tr>
<tr>
<td>#3</td>
<td>Z-Bond</td>
<td>2 CA101 dip</td>
</tr>
<tr>
<td>#4</td>
<td>Z-Bond</td>
<td>3 CA101 dip</td>
</tr>
<tr>
<td>#5</td>
<td>ZP-15E</td>
<td>Elastomer</td>
</tr>
<tr>
<td>#6</td>
<td>ZP-15E</td>
<td>Wax</td>
</tr>
<tr>
<td>#7</td>
<td>ZP-130</td>
<td>None</td>
</tr>
<tr>
<td>#8</td>
<td>ZP-130</td>
<td>Wax</td>
</tr>
<tr>
<td>#9</td>
<td>ZP-140</td>
<td>None</td>
</tr>
<tr>
<td>#10</td>
<td>ZP-140</td>
<td>Water</td>
</tr>
<tr>
<td>#11</td>
<td>Z-Cast</td>
<td>None</td>
</tr>
<tr>
<td>#12</td>
<td>Z-Cast</td>
<td>Baked</td>
</tr>
</tbody>
</table>

**Table 3-1:** Summary of example materials tested for simulation of bony and soft tissues in the prototype simulator.

A simple test object honeycomb plate shown in Fig. 3-2 was designed using a 3D software program (SolidWorks® 2012, Dassault Systèmes SolidWorks Corp., Waltham, MA) and printed using a 3D printer Z310 (Zcorp, Burlington, Ma). The honeycomb model provided qualitative evaluation of drilling purchase (solid rectangular base), cut-ability (hexagonal walls), and quantitative evaluation of material density (HU).
Fig 3-2 Honeycomb 3D printing test materials. Developed to test drilling purchase, cut-ability and density of various powder plaster materials post-processed according to details in Table 3-1.

Drilling was performed using a 4 mm diamond bur and 3 mm cutting bur with an otologic drill (Medtronic Xomed, Jacksonville FL), and cutting of the materials was tested with standard surgical equipment including Kerrison rongeurs (Karl Storz, Tuttlingen, Germany) and through-cutting instruments. The materials that most closely resembled native sinus tissue air cells as assessed by an expert sinus surgeon were selected.

Similarly, for the flexible materials simulating cartilaginous tissues of the turbinates and septum, a number of materials and post-processing techniques were considered. Post-processing was the critical step in achieving realistic tissue properties – yielding a plaster-based material that could be flexed and/or cut in a manner similar to soft-tissue. The turbinates and septum were printed using ZP-15 powder plaster and were post-processed in an elastomeric infiltrant to achieve the necessary flexibility. The flexibility was optimized according to the number of post-processing
“dips” and processing time of 30 minutes. In some cases, it appeared that the flexible and malleable structures had expanded during the post printing processing due to the process of dipping in the elastomeric infiltrant (e.g., a 5-10% expansion in the surface area of the anatomical structure). To assemble the components into the final phantom, these needed to be modified with the aid of surgical tools to reduce the width and length.

3.2.4 Accuracy and Reproducibility of Material Printing

The process was tested using a simple, low-fidelity design that incorporated a 10 mm diameter sphere created with 3D visualization software (Mimics v12.0, Materialise, Ann Arbor MI) as illustrated in Fig 3-3(a,b). This was done for both the firm bony structures (ZP-130 material) as well as the flexible structures (ZP-15 with elastomeric infiltrant). The actual diameter of the spheres was measured using digital calipers (Digimatic Mitutoyo Corporation, Japan), with 5 measurements repeated across 10 separate spheres, for a total 50 measurements. The mean diameter and standard deviation was calculated for each and compared to the design spec (10 mm) to evaluate accuracy and reproducibility.
Figure 3-3 (a) Ten ZP 15 and (b) ten ZP 130 spheres. Designed to test accuracy and reproducibility of the Z-310 printing process in bony and soft tissue structures, respectively.

3.2.5 Minimum Feature Size

A simple test object analogous to a line-pair pattern as illustrated in Fig 3-4 was designed using SolidWorks (SolidWorks® 2012, Dassault Systèmes SolidWorks Corp., Waltham, MA) to evaluate the minimum feature size that could be achieved with high-fidelity in the 3DP process. The test object presents a series of legs (a comb function) at a specified depth (6 mm) and with sequentially decreased distance between legs ($\Delta = 5, 4, 3, 2, 1, \text{ and } 0.5 \text{ mm}$, with corresponding
spatial frequency \( f = 1/\Delta \). The actual depth was measured in each interval, and the ratio of actual to specified depth is the “modulation” (i.e., the fidelity) with which the process replicates the design.

![Image of feature pattern to test minimum feature size.](image)

**Fig 3-4 Feature pattern to test minimum feature size.** A depth of 6mm was designed between each leg, with the actual depth corresponding to feature “modulation.”

### 3.2.6 Qualitative Analysis of Drilling Purchase

The density of the ZP-130 powder can be varied depending on the percentage of mixture with the CA101 cyanoacrolate infiltrant. Test samples (solid blocks) were created that varied the percentage infiltrant (10%, 20%, 50%, and 70%) to determine qualitatively the mixture that most closely resembled the tactile feedback of drilling through the bone of the skull base. Both 3 mm cutting burrs and 4 mm diamond burrs on an otologic drill were used by an expert surgeon.
3.2.7 Accuracy and Qualitative Analysis of Air Cell Walls and “Membranes”

Test objects were also produced using (SolidWorks® 2012, Dassault Systèmes SolidWorks Corp., Waltham, MA) to determine the minimum thickness soft-tissue-simulating “membrane” that could be printed and the qualitative tactile / cut-ability features associated with the varied thicknesses. The impetus for these tests was to examine material characteristics associated with membrane like qualities of the peri-orbital tissue (lamina papyracea), the medial wall of the maxillary sinus, and the bony walls separating ethmoidal air cells (typically less than 1mm in thickness). A test object is illustrated in Fig. 3-5 where each circular region corresponds to a specified thickness (5, 4, 3, 2, 1, and 0.5 mm). First, the accuracy of the printed membrane thickness was measured using digital calipers, repeated 5 times at random points within each circular region. Second, the same membrane samples were evaluated by an expert surgeon in qualitative tests of surgical cutting and punching maneuvers. This was accomplished by using surgical instruments to punch and cut through the circular test region at each specified thickness until the membrane was removed in its entirety.
**Figure 3-5 Flexible Zp-15 test plate.** Developed to test the accuracy and minimum thickness of soft-tissue “membrane” simulating material, with membranes as thin as 0.5mm.

### 3.2.8 Statistics

Data was analyzed using descriptive statistics and comparison of means were made using student’s t-test for paired samples. Statistical software used for data analysis was SAS 9.13 (SAS statistical software, Cary North Carolina)

### 3.3 Results
3.3.1 Material Development

<table>
<thead>
<tr>
<th>Module #</th>
<th>Material</th>
<th>Post-Processing</th>
<th>Qualitative Assessment of Tissue Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Drilling</td>
</tr>
<tr>
<td>#1</td>
<td>Polycarbonate</td>
<td>None</td>
<td>n/a</td>
</tr>
<tr>
<td>#2</td>
<td>Z-Bond</td>
<td>1 CA101 dip</td>
<td>Modest</td>
</tr>
<tr>
<td>#3</td>
<td>Z-Bond</td>
<td>2 CA101 dip</td>
<td>Modest</td>
</tr>
<tr>
<td>#4</td>
<td>Z-Bond</td>
<td>3 CA101 dip</td>
<td>Modest</td>
</tr>
<tr>
<td>#5</td>
<td>ZP-15E</td>
<td>Elastomer</td>
<td>Poor</td>
</tr>
<tr>
<td>#6</td>
<td>ZP-15E</td>
<td>Wax</td>
<td>Poor</td>
</tr>
<tr>
<td>#7</td>
<td>ZP-130</td>
<td>None</td>
<td>Good</td>
</tr>
<tr>
<td>#8</td>
<td>ZP-130</td>
<td>Wax</td>
<td>Excellent</td>
</tr>
<tr>
<td>#9</td>
<td>ZP-140</td>
<td>None</td>
<td>Good</td>
</tr>
<tr>
<td>#10</td>
<td>ZP-140</td>
<td>Water</td>
<td>Good</td>
</tr>
<tr>
<td>#11</td>
<td>Z-Cast</td>
<td>None</td>
<td>Poor</td>
</tr>
<tr>
<td>#12</td>
<td>Z-Cast</td>
<td>Baked</td>
<td>Poor</td>
</tr>
</tbody>
</table>

Table 3-2 Summary of Qualitative Assessment of Example materials. Tested for simulation of bony and soft tissues.

The materials that most closely resembled native sinus tissue air cells as assessed by an expert sinus surgeon were selected following blinded surgical manipulation of the 12 honeycomb plates. For the bony structures, the chosen material was a ZP-130 powder plaster (ZCorp, Burlington, Ma) with post-processing by 15-minute infiltration by CA101 cyanoacrolate. For the turbinates and septum, the optimal material was the ZP-15 powder plaster that was post-processed in an elastomeric infiltrant to achieve the necessary flexibility. The summary of the qualitative assessment of all 12 honeycomb plates is shown in table 3-2.

3.3.2 Reproducibility and Accuracy

The mean diameter and standard deviation was calculated for each and compared to the design spec (10 mm) to evaluate accuracy and reproducibility. The mean diameter and standard
deviation for the bony spheres (ZP-130) was 9.97 ± 0.05 mm. For the soft-tissue spheres (ZP-15), the mean diameter and standard deviation was 10.02 ± 0.07 mm. Accuracy with respect to the specified (10 mm) diameter was therefore within 3%. The mean relative difference was therefore 0.3% and 0.2% respectively for the bony and soft tissue spheres. There was no statistically significant difference in mean diameter when comparing the ZP-130 spheres with the ZP-15 spheres (p = 0.16).

3.3.3 Modulation and Feature Size

As the separation begins to decrease beyond 3mm ($f \sim 0.33 \text{ mm}^{-1}$) the modulation begins to significantly drop off. The measured feature modulation is plotted vs. spatial frequency in Fig. 3-6. As the separation approaches 0.5mm ($f \sim 2 \text{ mm}^{-1}$) the modulation drops to ~0.20. These measurements evaluate the approximate threshold for separation of two distinct objects in the phantom, in effect the spatial resolution of the 3DP process, an important consideration, since critical anatomical structures are often separated by as little as 1-2 mm.
Figure 3-6 Modulation in 3DP feature size measured as a function of frequency (separation). Modulation dramatically decreases beyond separation of ~3mm ($f \sim 0.33$ mm$^{-1}$). The results suggest that, depending on feature size, the resolution between distinct objects degrades for features less than ~3mm.

3.3.4 Qualitative Analysis of Drilling Purchase

A blinded expert surgeon found no discernable difference qualitatively between the 10%, 20%, 50% and 70% CA 101 cyanoacrolate infiltrant of the ZP – 130 test sample blocks. All samples had similar tactile/drilling purchase characteristics that resembled human cortical bone.
3.3.5 Qualitative Analysis and Accuracy of Air Cells and Membranes

At the largest thicknesses, the membrane thickness was accurate to within ~5%, but accuracy degraded considerably at 1 mm or below for the ZP-130 (bone-simulation) and at 3 mm or below for the ZP-15 (soft-tissue simulation). The mean thickness and standard deviation with accuracy is reported in Table 3-3 for all specified thicknesses from 0.5mm – 6mm. There was a significant difference between the ZP-130 membrane and the ZP-15 membrane in average thickness (p = 0.02). Second, the same membrane samples were evaluated by an expert surgeon in qualitative tests of surgical cutting and punching maneuvers. Membranes thicker than ~3 mm [for both the ZP-15 (flexible) and ZP-130 (firm) materials] were not able to be cut or punched with traditional surgical instrumentation and therefore, they were not able to be assessed.

Based on assessment by a non-blinded expert, the 0.5 mm and 1.0 mm thick ZP-130 and ZP-15 membranes satisfactorily resembled the true resectability of the periorbital wall, medial maxillary wall, and ethmoid sinus air cells; however, 0.5mm was the optimal thickness.
Table 3-3 Mean depth and standard deviation of ZP-130 and ZP-15 test samples of membrane thickness.

Although accuracy of 0.5 and 1mm membrane thickness was poor, these yielded optimal qualitative drilling/cutting properties comparable to peri-orbital tissue and medial maxillary or ethmoidal cell membranes, respectively.

3.3.6 Prototype High-fidelity Sinus Skull Base Simulator

The resulting sinus and skull base simulator prototype is illustrated in Fig 3-7. The bone module 3-7 (a) exhibits realistic anatomical features with drilling purchase and density approximate to bone. Note the realistic architecture of the frontal, ethmoid, and sphenoid sinuses. The soft tissue modules [Fig. 3-7(b)] exhibit flexibility and cut-ability approximate to cartilage. The simulated septum, middle turbinates, and lower turbinates snap into the sinus cavity, and the phantom is supported within an anthropomorphic head phantom as in Fig. 3-8 (b). The setup of the sinus/skull base simulator during an assessment for surgical task performance is depicted in Fig 3-8.
Figure 3-7 Illustration of high-fidelity sinus and skull base prototype for surgical simulation. (a) rigid bone like material and (b) flexible material. CT images (c) illustrate the anatomical realism of the model and confirm tissue density properties.
Figure 3-8 Experimental setup for surgical task performance. (a) photograph of surgeon performing endoscopic surgical tasks, (b) close-up view of surgical tools and endoscope inserted in the nasal cavity, (c) endoscopic video image in the simulated paranasal sinuses.

3.4 Discussion

Rapid prototyping is based on the digital conversion of image sets to a three dimensional model. The images are acquired on conventional hospital scanners (e.g., CT, MRI) from numerous patient anatomic regions. The primary uses for rapid prototyping in the medical community has revolved around pre-operative planning and medical modeling in a number of different surgical specialties. The application of rapid prototyping to medical modeling can be divided into three distinct stages: data acquisition, image processing and model production. Image processing is often the most time consuming stage of the process. Highly skilled individuals with a comprehensive understanding of both the anatomical/clinical issues and the technical details of RP are required to produce models that are valid and relevant. The process of segmentation of the images acquired by computed tomography involves three distinct steps. The first step involves intensity thresholding which can be either automatic or semi-automatic whereby the
software platform will segment the corresponding anatomy based on the Hounsfield units.
Following intensity thresholding the technique of region growing is employed which allows for anatomical structures that are in close proximity and share similar imaging characteristics to be included in the desired segment. The final step which is the most important and time consuming, involves manual segmentation in all three orthogonal planes. This stage is the most critical and labor intensive element of creating an anatomically accurate model. As the model increases in complexity the third stage in segmentation, manual contouring, increases in duration considerably.

The use of RP technology for the purposes of surgical simulators that can be used for training is a new concept. RP has been described and shown to assist in planning of individual patient cases. Muller et al. investigated the utility of RP in the models of the skull both in a craniofacial and neurosurgical practice. They looked at 52 patients that were grouped into one of the three different surgical procedures: 1) corrective cranioplasty 2) reconstructive cranioplasty or 3) planning of difficult skull base approaches. They were able to show that operative time and surgical errors were reduced in the corrective cranioplasty group. In addition, the RP models were able to provide a better understanding of the anatomy, increase intraoperative accuracy, support accurate fabrication of implants, facilitate pre-surgical simulation and improve education of surgical trainees. Others have shown the utility of RP for soft tissue components. Binder et al. looked at 24 mitral valve specimens from transesophageal echocardiography that were printed with RP technology. They found that there was good agreement between the phantoms and the replicas in volume and maximal dimensions. In addition, visual and tactile examination of the mitral valve phantoms by two blinded observers allowed correct depiction of mitral valve anatomy and pathology in all cases.
In our study, we have been able to demonstrate after qualitative analysis of numerous bone and soft tissue products created by 3DP, that there is one printing material, ZP-130 with CA101 infiltrant, that most closely resembles the drilling purchase and feel of bone found within the paranasal sinus & skull base, specifically the clival region. When evaluating the soft tissue component that would represent the cartilaginous and soft bony structures of the septum and turbinates, the ZP-15 with elastomeric application was the closest match. In order to create a simulator that has strong face validity one must be able to create realistic anatomical structures that closely resemble the tactile and haptic feedback of native human anatomy.

Qualitative analysis by an expert surgeon was also used to determine the optimal thickness of membranes and cell walls. Numerous anatomical areas within the sinus and skull base region such as periorbita, lamina papyracea, ethmoidal air cell wall and medial maxillary wall act as a membrane like structure. In order to create a model that would be receptive to traditional surgical instrumentation we determined the optimal and threshold thickness for both the bony walls and soft tissue membranes. Based on the data obtained in this study, 2 mm appears to be the limit of thickness for structures that require surgical resection. Therefore, sinus models with wall thickness greater than 2 mm are not expected to be realistically receptive to traditional surgical cutting and punching instrumentation (e.g., Kerrison rongeurs, Through-Cutting forceps, etc.). This will be a useful benchmark for development of future simulation models with 3DP in the head and neck regions.

Rapid prototyping is a sequential, multi step process. Errors in production of prototypes, or as in this study, models for surgical simulation can occur at any step in the production process. This can potentially affect shape, dimensions and anatomic details. The error can occur in acquisition of the images with CT scans, processing and manual manipulation of the images with 3D
visualization software or in the fabrication and post processing stages. Silva and colleagues compared the accuracy of 3DP with SLS in the production of a dry skull rapid prototype. A number of predefined anatomic dimensions and landmarks were measured both in the dry skull and the prototypes with digital calipers and measurements were repeated 20 times. They concluded that both 3DP and SLS techniques produced acceptable precision and they felt that they could be used as guides for most maxillo-facial surgery. However, 3DP had a slightly higher mean error of 2.67% vs 2.10% for SLS. They also concluded that the models produced acceptable anatomic details except for thin bones, small foramina and acute bony projections. They did not describe the specific dimensions, measurements or cut off values for these structures. An important question to ask is how much error is acceptable in the prototyping process. Asaumi et al looked at rapid prototype models in the management of coronoid hyperplasia and felt that 2% variation would be acceptable and would not significantly impact surgical management. The amount of variation that can be tolerated is directly related to the task for which the prototype is being used. If the goal of the prototype is to communicate with patients and their families, surgical education, or for simulation of large defects a larger variation can be tolerated.

The accuracy and reproducibility was evaluated during our production process and when looking at reproducibility we found there was a less than 0.3% mean relative difference in the production of 10mm spheres using the 3DP process. This compares favorably to other studies done by Choi and colleagues that showed a mean relative difference of 0.56%. However, their measurements included 16 linear measurements over different spots on the skull.

We performed a calculation to determine the modulation of the rapid prototyping process. This to the best of our knowledge has not been reported on previously in the rapid prototyping
literature. The modulation represents the fidelity of the printing process, or in simpler terms the ability to print two objects that are close to each other without the separation between them being compromised. This is especially important with respect to design of a simulator in the head and neck region and specifically with respect to the paranasal sinuses and skull base. Many anatomical structures are separated by only a few millimeters such as the posterior aspect of uncinate process and the anterior aspect of the ethmoid bulla. The space between these two anatomical structures forms the hiatus semilunaris. This is a surgical access point for an uncinectomy and antrostomy. Therefore understanding the minimal distance two items can be separated will be important in simulator design.

The modulation of the 3DP process was maintained as long as the separation between two objects was greater than 3mm. Once objects were separated by less than 3mm the fidelity of the 3DP process began to drop off significantly down to a level of 20% of expected depth at 0.5mm. This decrease in fidelity can be attributed to one of two issues in the 3DP process. The first issue is a technical limitation of the printing process itself. During the printing process the printer has a reservoir for the polymeric powder, a build tray that moves down, a roller to distribute and evenly spread the layer of powder, and a print head that distributes a binding material. First, the roller moves over the build tray and evenly spreads a uniform layer of powder; the print head moves in the X and Y axes and releases a jet of binder onto the powder; and the binder fuses the powder. After that, the platform moves down; another layer of powder is deposited and receives the jet of binder. This second layer fuses and adheres to the previous layer, and the process is repeated. The second issue occurs when the model is ready, the remaining loose powder is aspirated from the surface. As this process does not confer great resistance, models have to undergo a finishing process that consists of addition of infiltrating materials, such as cyanoacrylate. This infiltration process has the potential to introduce further loss of fidelity
between two objects that would otherwise be separated.

In addition to determining that the cyanoacrylate can potentially lead to a loss of fidelity we also explored whether the concentration of the cyanoacrylate used would change the properties of the bone structure that was printed. This was evaluated qualitatively by an expert surgeon and the concentration of the cyanoacrylate had no bearing on the bone quality as tested with a surgical drill. This to our knowledge has not been previously reported.

The final analysis involved determining the membrane and bone thickness that could be printed with accuracy and precision. This has important implications when creating anatomical structures such as the medial orbital wall and thin bone overlying the skull base that separates the intracranial contents from the sinonasal tract. We found that once the bone thickness in the ZP-130 material fell below 2mm and the soft tissue thickness fell below 5mm, the accuracy of the membrane and wall thickness fell below the 2% cut off that has been previously discussed. The difference in the bone thickness and membrane thickness can be directly attributed to the elastomeric infiltrant and its effect on the thickness of the soft tissue components. This has also been commented on by previous authors \(^{57,105}\), however, the quantification of the magnitude of inaccuracy in the 3DP process based on type of infiltrant has not been previously reported. The implication of this finding does not prohibit the use of elastomeric infiltrant to create soft tissue structures, however, it does provide a rough guide in order to anticipate the expansion of anatomical units during production.

One of the challenges in evaluation of surgical aptitude within the paranasal sinuses and skull base is the lack of models and simulators that are realistic, reproducible, and appropriate for a broad spectrum of normal anatomy, anatomical variation and specific disease models. The sinus/skull base simulator described in this paper will hopefully provide a platform for this. The
sinus/skull base simulator is based upon high-fidelity models created from cadaveric CT images and realized via 3DP and rapid prototyping. Iterative testing and development of novel material formulations simulating bony and soft tissue structures is essential to creating realistic models that can be used for simulation. This approach provides the capacity to conduct patient specific simulation prior to embarking on an actual operation. The ability to do so in realistic physical models (as opposed to conventional digital simulation) could offer significant improvement in preparation for the most challenging surgical cases. The creation of realistic surgical models also provides a valuable platform for critically analyzing the development of novel technologies and new techniques of evaluating surgical aptitude. The ability to create identical tasks for a given number of surgeons permits quantitative evaluation of improvements in surgical efficacy and safety. One question that remains to be answered is whether this simulator will withstand the rigors of simulator validation so that it can be incorporated in medical education of surgical trainees.
Chapter 4

4 Validation of a Sinus/Skull Base Surgery Simulator developed with Three-Dimensional Printing

4.1 Introduction

Simulation has been defined as “the imitation of the operation of a real world process or system over time.” The notion of simulation to improve skill and aptitude at performing a given task has been theorized to be present for centuries. The arena of simulation that has been most often tied to the field of medicine is aviation simulation. The ability to recreate stressful and high risk medical environments with simulated environments allows for the creation of a safe and supportive educational climate. The features of simulation that make it attractive for application in medical education include: standardization, availability and repeatability. Availability allows for integration in complex curriculums with many time constraints. Repeatability allows for the ability to select findings/conditions/situations and complications that are relevant for a given trainee. These attributes allow the student to train for both clinical skills, in addition to, high stakes certification examinations.

Simulation for medical education has been categorized using a number of different schemes. One of the most comprehensive classifications is based on five distinct categories that include: verbal, standardized patients, part-task trainers, computer patient and electronic patient. Verbal simulation is essentially role-playing. Standardized patients (SP’s) are used in a variety of roles, including: history taking, physical examination skills, communication and professionalism.
SP’s can be used either to educate or evaluate trainees. Part-task trainers can range from simple anatomic models to represent various regions of the human body, to complex modern surgical task trainers that are used to simulate entire operations. Part-task trainers are widely employed in the education of novice surgical trainees. Computer based patients are a low-cost alternative to traditional SP’s and can either be software based or web-based. The final category is the electronic or digital patient. These “electronic” patients are either mannequin based or virtual reality systems. The clinical environment that is being simulated in conjunction with the electronic patient is important for simulator effectiveness. The simulator that we will be evaluating would fall under the category of part-task trainer.

Within the field of Otolaryngology-Head and Neck Surgery (OHNS) there are two specific anatomic regions and procedures that have been most studied with respect to simulator validation and development. These are endoscopic sinus surgery (ESS) and temporal bone surgery. The first and most widely studied and validated simulator for endoscopic sinus surgery is the Endoscopic Sinus Surgery Simulator (ES3). This simulator was developed in 1997 and was created through a partnership between Lockheed Martin and the Universities of Washington, Ohio State and Ohio Supercomputer Center. The impressive validation of the ES3 has allowed it to be considered one of the most robust virtual simulators that has been developed. It has shown both construct validity, in its ability to distinguish between various levels of surgical experience, and predictive validity. Predictive validity is the least reported of all the types of validation. A recent review by Van Nortwick and colleagues showed that it is reported less than 5% of the time in simulator validation publications, despite it being considered the most valuable of all the types of validation methods. Fried was able to show the predictive validity of the ES3 with junior residents who had been exposed to training on the ES3 exhibiting improved skills in instrument manipulation, shorter operative times, reduced technical errors and more
confidence in real endoscopic sinus surgery when compared to junior residents who did not receive simulator training. There are a number of limitations to the implementation of the ES3 in surgical training programs despite its track record. The initial major obstacle was cost, however, since the ES3 is no longer in production and there are only a small number of systems that are in existence, accessibility is now the major barrier.

In addition to part-task trainers and virtual reality simulation, others have reported on their experience with high-fidelity medical simulation within OHNS. This falls under the electronic patient category with a simulated environment. Participants in this study, which included PGY-3 OHNS residents, fellows, anaesthesia residents and fellows and nursing staff, were exposed to 3 simulated medical crises that centered on the airway and anesthesia issues. The authors concluded that this form of simulation can effectively teach clinical decision making skills and teamwork concurrently to members of an OR team.

Traditional methods of surgical simulation that have been employed in OHNS surgical curriculum have relied upon cadaveric specimens from the head and neck for both sinus surgery and temporal bone surgery. Cadaveric dissection still remains the gold standard to which all new novel simulators, whether they be partial task-trainers or electrical patients (VR systems), are compared. The limitations of traditional cadaveric dissection include potential for infection risk (HIV, Hepatitis B, Hepatitis C), physical tissue properties from the fixation technique that was employed, availability of specimens and ethical concerns in certain parts of the world. These limitations have driven the desire to develop new forms of part-task trainer simulation to teach surgical skills to novice trainees. One of the recent technologic advancements that has allowed the creation of anatomic models involves three-dimensional printing which is an additive form of manufacturing. We have previously described the design and material selection in the
development of a sinus and skull base simulator based on 3DP for the purposes of training novice surgical residents.  

As surgical simulation continues to be incorporated into all surgical residency programs, there is a need for cost-effective validated simulators to teach pre-defined skills away from the real-patient interaction. In OHNS there is a paucity of simulation devices available for acquisition of endoscopic sinus and skull base surgery core skills. We have developed a high-fidelity sinus and skull base simulator with the aid of 3DP. The aims of this study are to:

1) Determine face, construct and concurrent validity of our Sinus/Skull Base Simulator.

2) Evaluate novel surgical tool kinematics during simulation.

4.2 Materials and Methods

4.2.1 Subject Recruitment

After obtaining approval from our institutional review board, twenty eight subjects were enrolled in the study. All participants were Canadian trained otolaryngology head and neck surgery (OHNS) residents, fellows or faculty. Residents were from post graduate years one through five (PGY 1 through PGY 5), taking part in an annual endoscopic sinus surgery course. Fellows were OHNS trained physicians seeking further sub-specialty training in rhinology and skull base surgery. Faculty members were OHNS faculty with a sub-specialty interest in rhinology and
skull base surgery. Faculty and fellows were supervising and assisting in an Endoscopic Sinus/Skull Base Surgery course that was being held at the University of Toronto Surgical Skills Centre. Subjects were stratified into two groups based on level of experience: novice and expert. Demographic data including age, handedness, length of training, number of endoscopic sinus surgeries previously performed and prior exposure to simulators was recorded.

4.2.2 Simulator Assembly

Thirty two simulators were produced using a 3D Spectrum Z510 printer (Zcorp, Burlington, Ma). The cadaveric head was scanned on a GE Lightspeed 16-slice CT scanner (GE Healthcare, Milwaukee, WI) with isotropic (1 mm) voxel size. Digital processing of the images was performed using 3D visualization software (Mimics v12.0, Materialise, Ann Arbor MI). An "stl" file was generated from a 3D surface model of the CT imaging data set and printed with ZP-131 material and ZBond-101 infiltrant. For soft tissue components (turbinate and septum), a 2-part silicone compound (DragonSkin 30) was used. A negative mold was created from the 3D surface model and printed using the same process. After printing, the silicone mixture was poured into the mold and de-gassed for approximately 15-20 minutes. The de-gassing was repeated until there were no more air bubbles visible. The mold was placed into an oven and heated to ~150°F for 1 hour. Afterwards, the silicone model was separated from the mold. Once the soft tissue components and bony components were printed and treated, they were manually assembled and placed in the mannequin (Figures 4-1, 4-2). The cadaveric head was selected based on its normative endoscopic anatomy and was reused by all subjects. The tasks were not ablative and therefore, allowed for repeated use.
Figure 4-1 Assembly of Sinus/Skull Base Simulator The simulators are assembled with the left and right side as separate components in order to facilitate placement of the soft tissue structures that must be placed manually.
Figure 4-2 Simulator Assembled. Once the simulator is assembled it is placed in the foam head ring and the optical tracking navigation is fixated on the forehead with adhesive tape. This is the view the subjects had prior to surgical simulation.

4.2.3 Surgical Tasks

The subjects performed two surgical tasks in the cadaver model and five surgical tasks in the sinus/skull base simulator. The two surgical tasks in the cadaver consisted of three pass endoscopy using a zero degree endoscope (Karl Storz, Tuttingen, Germany) and placement of a quilting suture in the nasal septum. The three passes performed involved nasopharynx identification, followed by anterior sphenoid face/superior meatus localization and middle meatus endoscopy. The quilting suture involved using the endoscope to guide placement of a 4/0
nylon suture through the anterior nasal septum. The five surgical tasks in the simulator included: three pass endoscopy, a quilting suture of the nasal septum, left middle turbinate resection, bilateral wide sphenoidotomies with the aid of a high speed sinus drill (Medtronic Xomed, Jacksonville, Fl) and binarial identification of the Sella Turcica.

4.2.4 Assessment Tools

To rate the subjects performance on both the cadaver and simulator, five expert assessors used a two part assessment tool that had been developed at Johns Hopkins University by Lin and colleagues. It consists of a task specific checklist portion and a global rating portion. The checklist section comprises procedure-specific tasks that are equally weighted that allows for an evaluator to provide formative feedback during each step of the procedure. Certain portions of the task specific checklist and global checklist had to be omitted, as they could not be evaluated by video assessment. The simulator global rating (SIM-GRS) instrument consists of a 6-item form with a 5-point Likert scale linked to concrete descriptors (modified from Reznick). The six items on the SIM-GRS are described in Table 4-1.
1) Use of Endoscopes
2) Instrument Handling
3) Respect for Tissue
4) Time and Motion
5) Flow of Operation
6) Overall Surgical Performance

Table 4-1 Simulator Global Rating Scale. Six items that were evaluated by expert assessors on the video analysis of global performance on the simulator.

The simulator task specific scale (SIM-TSS) instrument consists of seven items with a 5 point Likert scale, also linked to concrete descriptors (Table 4-2).

<table>
<thead>
<tr>
<th>SIM-TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: Sinus Endoscopy</td>
</tr>
<tr>
<td>a) Inferior Pass (Nasopharynx)</td>
</tr>
<tr>
<td>b) Superior Pass (Sphenoid Face)</td>
</tr>
<tr>
<td>c) Intermediate Pass (Middle Meatus)</td>
</tr>
<tr>
<td>Task 2: Septum Suture</td>
</tr>
<tr>
<td>Task 3: Middle Turbinate Resection</td>
</tr>
<tr>
<td>Task 4: Bilateral Sphenoidotomy</td>
</tr>
<tr>
<td>Task 5: Binarial Identification of Sella Turcica</td>
</tr>
</tbody>
</table>

Table 4-2 Simulator Task Specific Scale (SIM-TSS). Seven items among the five tasks that were evaluated during video assessment of the subjects on the simulator.
The cadaver global rating scale (Cad-GRS) was the same six item questionnaire as used on the simulator (SIM-GRS). The cadaver task specific checklist (Cad-TSS) was further modified and included four items under the heading of two tasks (Table 4-3).

<table>
<thead>
<tr>
<th>Cad-TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1: Sinus Endoscopy</td>
</tr>
<tr>
<td>a) Inferior Pass (Nasopharynx)</td>
</tr>
<tr>
<td>b) Superior Pass (Sphenoid Face)</td>
</tr>
<tr>
<td>c) Intermediate Pass (Middle Meatus)</td>
</tr>
<tr>
<td>Task 2: Septum Suture</td>
</tr>
</tbody>
</table>

Table 4-3 Cadaver Task Specific Checklist (Cad-TSS). Four items among the 2 tasks that were evaluated during video assessment of the subjects on the cadaver.

Furthermore, the degree of realism (face validity) of the simulator was assessed using an eleven item realism questionnaire on a seven-point Likert scale (‘‘1 – absolutely not realistic’’, ‘‘2 – not realistic’’, ‘‘3 – somewhat not realistic’’, ‘‘4 – undecided’’, ‘‘5 – somewhat realistic’’, ‘‘6 – realistic’’, and ‘‘7 – absolutely realistic’’). Many types of questionnaires have been used in previous studies. We chose the one that fits the best and was easiest to adapt to our simulator. The face validity questionnaire was administered to all 28 subjects partaking in the study and the qualities that were assessed can be seen in Table 4-4.
### Table 4-4 Level of Realism

<table>
<thead>
<tr>
<th>Level of Realism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) First Impression</td>
</tr>
<tr>
<td>2) Design of the Simulator</td>
</tr>
<tr>
<td>3) Mannequin or Head Phantom</td>
</tr>
<tr>
<td>4) Haptic Feedback</td>
</tr>
<tr>
<td>5) Movement with Optics</td>
</tr>
<tr>
<td>6) Quality of Simulated Tissue</td>
</tr>
<tr>
<td>7) Cut-ability &amp; Flexibility of Cartilage like tissue</td>
</tr>
<tr>
<td>8) Drilling Purchase and Density of Bone Like Tissue</td>
</tr>
<tr>
<td>9) Endoscope Adaptation</td>
</tr>
<tr>
<td>10) Navigation with Endoscope</td>
</tr>
<tr>
<td>11) Overall Impression</td>
</tr>
</tbody>
</table>

**Table 4-4 Level of Realism** Eleven questions that were asked of the participants to determine face validity

### 4.2.5 Instrument Tracking/Hand Motion Capture & Analysis

The University Health Network Guided Therapeutics (GTx) program has developed a custom image-guided surgery (IGS) system to provide “always-on” localization of surgical instruments (e.g., endoscope, drill, pointer) and real-time feedback for the surgeon. The system is denoted “Localized Intraoperative Virtual Endoscopy”, or LIVE-IGS, to reflect this paradigm. The main components of LIVE-IGS consist of a mobile C-Arm for intraoperative cone-beam CT, real-time 3D tracking devices (infrared/electromagnetic), a high-performance computer, LCD monitor and in-house navigation and visualization software (“X-Eyes”). X-Eyes was developed using the Image-Guided Surgery Toolkit, a well-known medical software library. The software platform provides multi-modality image display (e.g. CT, MRI, cone-beam CT), real-time tracking of surgical instruments, streamed endoscopic video, and overlays of surgical planning data.

Figure 4-3 shows the system implemented during our simulator validation study. An infrared (IR) stereoscopic camera (Polaris Spectra, NDI, Waterloo, Canada) is used to track surgical tools
mounted with tracker clamps containing IR reflective spheres. An electromagnetic (EM) table-
top field generator (Aurora, NDI, Waterloo, Canada) is used to track EM sensors attached to the
hands of the surgeon. Additional IR and EM reference markers are also attached to the phantom
to allow head movement with continued registration during the procedure. The X-Eyes
navigation software is capable of recording a near real-time (~20 frames per second) stream of
3D position and orientation from two IR-tracked surgical tools and two EM-tracked hands
simultaneously. This provides the ability to “playback” the movement of surgical tools and hands
over the course of the procedure. While surgical tool positions in an IGS system are typically
used to provide prospective navigation during the procedure, recording positional data enables
retrospective analysis of the characteristics of tool motion. In the context of this study, the
interest is in exploring correlations between this tool motion and both the surgical procedure and
skill of the operator.

Kinematics is the study of the motion of objects, and provides quantitative metrics to describe
motion trajectories including velocity, acceleration, total path length, and time travelled. These
metrics provide an objective method for evaluating the motion of surgical tools and hands. For
this study, kinematic metrics were computed from the time sequence of 3D positions recorded by
X-Eyes using MATLAB software (Mathworks, Natick, Mass). The following kinematic metrics
were included: time (s), tool separation (mm), path length (mm), jitter (mm/s3) and compactness
(mm).
4.2.6 Video Capture and Editing

The surgeries both on the cadaveric model and the sinus/skull base simulator were video recorded and edited by the primary author on a Macintosh MacBook Pro (Version 10.7.5; Apple Inc., Cupertino, Ca) using Elgato video capture (Version 1.0.1; Elgato Systems, San Francisco, Ca). The surgeries were edited to remove identifying information, including sound and other possible background identifiers in order for the identities of the subjects to remain unknown to the expert reviewers. The Macintosh software application iMovie (Version 9.0.9; Apple Inc., Cupertino, Ca) was used to edit the full length videos down to five separate tasks for the sinus/skull base simulator portion of the trial and two separate tasks for the cadaveric portion. This was done to facilitate and simplify the evaluation by the expert video reviewers. Each of the videos was labeled with a number from one to twenty eight from a random number generator in order for the expert reviewers to be able to track their evaluations while remaining blinded to their level of experience and identity.
4.2.7 Statistical Analysis

For data analysis we used SAS 9.13 (SAS statistical software, Cary North Carolina). The statistical analysis for the level of realism data used the Mann-Whitney U test for two group comparison. This test was chosen, as the distribution of our responses was not normally distributed. Analysis of the instrument tracking data used a repeated measures mixed model to compare outcomes between predictor categories. The repeated factor was the subjects. Predictor categories were Surgical Experience Level (expert or novice), Task (1-5) on the simulator, and Surgical Tool type (drill, endoscope or pointer). Used in this way, mixed models are constructed and interpreted in much the same way as an ANOVA or a General Linear Model. Initial models included interactions between Surgical Experience Level and Task and between Surgical Experience Level and Tool. Non-significant interactions were removed sequentially until an interaction was significant or only the main effects remained in the model. None of the outcome variables in the analysis were normally distributed because many observations were skewed to the right of the mean. A natural logarithm of all variables was employed to address this.

For the analysis of the video data we also used a repeated measures mixed model. The primary interest in this study was to assess the effect of surgeon experience on each of the thirteen outcomes under two testing conditions – cadaver and simulator. A secondary objective was to assess possible effects of other covariates on the outcomes. Surgeon experience was initially defined as whether or not surgeons were Experts or Novices. The covariates were: surgeon age in years, number of previous cases, length of training defined as PGY 1-3, PGY 4-5, Fellows or Staff, whether or not surgeons had previous simulator experience and right or left handedness.
Prior to model building, the data were explored to determine which covariates could be incorporated into the models. Mixed models were then constructed to assess the effects of the predictors on the outcomes. Mixed models were used because they allowed for inclusion of repeated measures factors which were subjects and raters. Initial models included raters as a repeated measures factor because all subjects were rated by five raters. However, model residuals were not sufficiently normally-distributed and model results were not reliable. Instead, we calculated the mean outcome score for each subject across the five scores received by the five raters and rater was removed from the model as a factor. Residuals from these models based on the mean outcome scores, were more normally-distributed and the results more reliable. Models initially included 2-way interactions but in all cases, interactions were not significant and were removed from models. Post-hoc comparisons were made using the Tukey HSD approach to multiple comparisons at the 0.05 alpha level of significance.

To assess reliability of the global rating scale and task specific checklist on the simulator a two way random model was used. The interrater reliability was calculated using the intra-class correlation coefficient (ICC). Cronbach’s $\alpha$ was calculated to determine the inter-item reliability.

In a previous study, the score recorded by the Cad-GRS in novices was 0.48 (score 2.4 = 48%), assuming standard deviation of 0.2 (20%) and true difference of 0.33 (score 1.68 =33%). A two sample two sided t test will provide a sample size of 12 (6 novices and 6 advanced) to be able to reject the null hypothesis that the performance level between novices and expert surgeons are equal in the sinus surgical simulator phantom with power 80%. The Type I error probability associated with this test of this null hypothesis is 0.05.
4.3 Results

4.3.1 Demographics

There were fourteen subjects in the novice group and fourteen in the expert group. The mean age of the novice surgeons was 31, with a range of 27 to 37 years of age. The mean age for expert surgeons was 42, with a range of 34 to 74. Ten of the fourteen (72%) novice surgeons were PGY 4-5. However, eleven out of fourteen (79%) had performed fifty or fewer sinus operations and all novice surgeons performed fewer than 100 sinus operations. All of the expert surgeons (14/14) had performed greater than 100 sinus operations. Nine (32%) of the subjects had previously worked with simulators, 6 of the novices and 3 of the experts (Table 4-5). Only 1 of the subjects was left handed, all the others were right handed.
<table>
<thead>
<tr>
<th>Group</th>
<th>Length of Training</th>
<th>PGY 1-3</th>
<th>PGY 4-5</th>
<th>STAFF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice Surgeon</td>
<td></td>
<td>4</td>
<td>10</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Expert Surgeon</td>
<td></td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4</td>
<td>10</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>(Number of Cases)</th>
<th>&lt;10</th>
<th>10-50</th>
<th>50-100</th>
<th>&gt;100</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice Surgeon</td>
<td></td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Expert Surgeon</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Previous Simulator Experience</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice Surgeon</td>
<td></td>
<td>6</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Expert Surgeon</td>
<td></td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>9</td>
<td>19</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Age</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice Surgeon</td>
<td></td>
<td>14</td>
<td>31.1</td>
<td>2.5</td>
<td>27.0</td>
<td>37.0</td>
</tr>
<tr>
<td>Expert Surgeon</td>
<td></td>
<td>14</td>
<td>42.2</td>
<td>11.8</td>
<td>34.0</td>
<td>74.0</td>
</tr>
</tbody>
</table>

**Table 4-5** Demographic information comparison of surgeon group by age, previous simulator experience, length of training and surgical volumes.
4.3.2 Level of Realism

The opinions of the novice and expert groups were compared using the Mann-Whitney U test to screen for dissimilarities. High p values (p > 0.20) indicate that no significant difference of opinion was found between the expert and novice groups. Twenty seven (96%) of subjects felt that the drilling purchase of bone like tissue was “6-realistic” or “7-absolutely realistic”. Twenty seven (96%) of the respondents also had a favorable evaluation of their first impression with the simulator, rating it as “6-realistic” or “7-absolutely realistic”. Six (21%) of the subjects felt “4-undecided” about the cut-ability of the tissues. Four (14%) felt that the quality of the soft tissue was “3-somewhat not realistic”. When evaluating overall impression with the simulator, none of the respondents felt that the simulator was “1-absolutely not realistic”, “2-not realistic” or “3-somewhat unrealistic”. The highest scores were for drilling purchase of bone (6.0 +/- 0.9) and navigation of endoscope (5.9 +/- 1.0). Lowest scores for the simulator were for quality of soft tissue (4.9 +/-1.3) and cut-ability of the cartilage like tissue (4.6 +/- 1.3)
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Impression</td>
<td>5.3</td>
<td>0.8</td>
<td>5.0</td>
<td>5.4</td>
<td>0.5</td>
<td>5.0</td>
<td>5.1</td>
<td>1.1</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Design of Simulator</td>
<td>5.5</td>
<td>0.6</td>
<td>5.0</td>
<td>5.4</td>
<td>0.5</td>
<td>5.0</td>
<td>5.6</td>
<td>0.7</td>
<td>5.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Mannequin</td>
<td>5.2</td>
<td>0.8</td>
<td>5.0</td>
<td>5.1</td>
<td>0.6</td>
<td>5.0</td>
<td>5.4</td>
<td>1.0</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Haptic Feedback</td>
<td>5.5</td>
<td>1.1</td>
<td>5.0</td>
<td>5.7</td>
<td>1.0</td>
<td>6.0</td>
<td>5.3</td>
<td>1.1</td>
<td>5.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Movement/Optics</td>
<td>5.5</td>
<td>1.2</td>
<td>5.5</td>
<td>5.4</td>
<td>1.2</td>
<td>5.0</td>
<td>5.5</td>
<td>1.2</td>
<td>6.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Quality of Tissue</td>
<td>4.9</td>
<td>1.3</td>
<td>5.0</td>
<td>5.2</td>
<td>1.1</td>
<td>5.0</td>
<td>4.6</td>
<td>1.5</td>
<td>5.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Cut-ability Cartilage</td>
<td>4.6</td>
<td>1.3</td>
<td>5.0</td>
<td>4.2</td>
<td>1.4</td>
<td>4.0</td>
<td>5.0</td>
<td>1.2</td>
<td>5.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Drilling Bone</td>
<td>6.0</td>
<td>0.9</td>
<td>6.0</td>
<td>6.0</td>
<td>0.9</td>
<td>6.0</td>
<td>5.9</td>
<td>1.0</td>
<td>6.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Endoscope Adaptation</td>
<td>5.8</td>
<td>0.8</td>
<td>6.0</td>
<td>6.0</td>
<td>0.6</td>
<td>6.0</td>
<td>5.5</td>
<td>0.9</td>
<td>5.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Navigation Endoscope</td>
<td>5.9</td>
<td>1.0</td>
<td>6.0</td>
<td>5.8</td>
<td>1.0</td>
<td>6.0</td>
<td>5.9</td>
<td>1.1</td>
<td>6.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Overall Impression</td>
<td>5.4</td>
<td>0.7</td>
<td>5.0</td>
<td>5.4</td>
<td>0.5</td>
<td>5.0</td>
<td>5.3</td>
<td>0.8</td>
<td>5.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 4-6 Level of Realism  Summary of realism data by surgeon type and overall score.

4.3.3  OSATS Analysis

Five expert evaluators rated fourteen novice and fourteen experts. Each subject had two videos (cadaver & simulator) for a total of fifty six videos for each evaluator to score. Each one of the videos was rated with the GRS and TSS for a total of 112 evaluations/expert and 560 evaluations cumulatively. The average length of the videos on the cadaver was 4 minutes and on the simulator the average length was approximately 13 minutes once the editing of unnecessary footage was completed.
The overall results of the 13 models are summarized in Table 4-7. Surgeon experience (novice or expert) was a significant predictor of mean scores for all measures except Task 4 (bilateral sphenoid drilling) and Task 5 (binarial identification of the sellar face). There was no difference in scores between performance on the cadaver vs simulator except for Task 1b (three pass endoscopy, superior pass) and Q3 on GRS (respect for tissue). On these two items, respect for tissue and three pass Endoscopy, Cadaver scores were higher than simulator scores. Previous exposure to a simulator was a non-significant predictor of all the outcomes.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Surgeon Experience</th>
<th>Mode of Testing</th>
<th>Previous Simulator Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1a</td>
<td>Novice vs Expert</td>
<td>Cadaver vs Simulator</td>
<td></td>
</tr>
<tr>
<td>Task 1b</td>
<td>0.001*</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Task 1c</td>
<td>0.002*</td>
<td>0.02*</td>
<td>0.8</td>
</tr>
<tr>
<td>Task 2</td>
<td>0.01*</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Task 3</td>
<td>0.02*</td>
<td>0.1</td>
<td>0.6</td>
</tr>
<tr>
<td>Task 4</td>
<td>0.05*</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Task 5</td>
<td>0.08</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>GRS Q1</td>
<td>0.14</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>GRS Q2</td>
<td>0.004*</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td>GRS Q3</td>
<td>0.006*</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>GRS Q4</td>
<td>0.0005*</td>
<td>0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>GRS Q5</td>
<td>0.002*</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>GRS Q6</td>
<td>0.0009*</td>
<td>0.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 4-7
**Table 4-7 Summary of the results** (p-values) of 13 statistical models assessing the effect of surgeon experience (construct validity), mode of testing (concurrent validity) and previous simulator experience, on performance rating on 7 surgical tasks and 6 Global Rating Scores. Values with “*” indicate a significant difference.

![Box plot showing GRS Q1 scores for novices and experts.](image)

**Figure 4-4 Example of both construct validity and concurrent validity on one of 13 items tested.** The difference between novice and expert surgeons on GRS Q1 supports construct validation. The similarity in scores amongst surgeons on the cadaver and simulator supports concurrent validity.

Post-hoc comparisons were performed to compare mean outcomes between the novice surgeons (PGY 1-3 and PGY 4-5) and expert surgeons to see how the three groups differ. In order to decide how the three groups differ, one must perform “post-hoc” tests. There was no differences found in either the SIM-GRS or SIM TSS amongst the PGY 1-3 and PGY 4-5 novice surgeons (Figure 4-5)
Figure 4-5  Post-hoc analysis showing no difference between novice surgeons on one of the surgical tasks on the simulator (p=0.76) using Tukey HSD analysis.

4.3.4  Inter-rater Reliability

The inter-rater reliability was very strong between the 5 expert reviewers on all 7 tasks performed on the task specific scale (SIM-TSS). All correlations exhibited an intra-class correlation (ICC) of greater than 0.7, which is considered a substantial correlation and 4 of the tasks exhibited an ICC of greater than 0.8, which is an outstanding correlation according to Landis and Koch. On the SIM-GRS, 5 of the 6 questions exhibited a substantial correlation with an ICC of greater than 0.7. On question 3 the ICC was 0.637 which is also considered a substantial correlation (Table 4-8). In all cases, the correlations were highly statistically significant (P<0.0001).
The Intra Class Correlation (ICC) amongst five expert raters on the two instruments Sim-GRS & Sim-TSS.

### 4.3.5 Instrument Tracking Analysis

#### 4.3.5.1 Elapsed Time

The first result of this analysis was the test of the interaction between Surgical Experience Level and Task. The interaction was not statistically significant (p=0.76) which means that the way in which the two groups differed was similar on all tasks performed on the simulator. The next
question to be answered is whether surgeons differ in time to complete a task and if so by how much? These results show us that log time differs significantly both between Tasks (p<0.0001) and between Surgical Experience Level (p=0.02). Figure 4-6 shows that novices take longer than experts and task 4 (drilling the sphenoid opening) takes much longer than the other tasks.

![Figure 4-6 Time Analysis](image)

**Figure 4-6 Time Analysis.** Novices take longer than experts (p=0.02) and the drilling task takes the most amount of time.

### 4.3.5.2 Tool Separation

Tool Separation is the distance measured between the tip of the endoscope and the tip of the pointer on task 1 or the drill on task 4 and 5. The interaction between Surgical Experience Level
and Task was not statistically significant (P=0.35) which means that the way in which the two groups differed in tool separation was similar on all tasks. After removing the interaction, the final model tells us that surgeons did not differ on tool separation based on level of experience (p=0.67) and that tool separation differs by task (p<0.0001).

Figure 4-7 Tool Separation. Tool Separation did not differ significantly by Surgical Experience Level, however, it did by Task.

4.3.5.3 Path Length

This is the total distance travelled by both tools in mm over the duration of each of the five tasks performed. There was no difference in this measurement between Novice and Expert Surgeons either between tasks (interaction between surgeon and tasks: P=0.82) or between tools (interaction between surgeon and tool: P=0.38) or overall. Figure 4-8 shows measurements by
Surgical Experience Level and Task. Perhaps, not surprisingly, measurements were overall more variable between novice surgeons than between expert surgeons.

![Path Length Graph]

Figure 4-8 Path Length. Path length by Surgical Experience Level and measurement type (each surgeon’s measurement was averaged over the two instruments used).

4.3.5.4 Jitter (Motion Smoothness)

Jitter is defined as a change in acceleration and represents motion smoothness. Jitter did not differ between surgeon type (p=0.1) or tool (p=0.66) though it did differ by task (p<0.0001). Figure 4-9 shows this data.
Compactness is a novel measurement of three dimensional standard deviation of the tool trajectory over time (figure 4-10). This measurement differed between task and tool (p<0.0001) but not between expert and novice surgeons (p=0.67) (Figure 4-11).
Figure 4-10 Compactness. Left (purple, endoscope) and right hand (yellow, endoscopic scissors) motion during turbinate resection task. This subject (Novice 3) had much more compact motion with the right hand in comparison to the left (by a factor of 3), likely due to many changes of the endoscope position during resection. The compactness metric is related to the size of ellipsoid shown which encapsulates the two hand positions.
Figure 4-11 Compactness by Surgical Experience Level and Task (each surgeon’s measurement was averaged over the two instruments used).

4.4 Discussion

This study demonstrates validity of the sinus/skull base simulator that was developed with the aid of 3DP. Validity was achieved in three domains: face, construct and concurrent validity. The first step in the process of validation is face validity, which is essentially a subjective evaluation of the simulator based on inspection and use. Despite the process being subjective, use of a systematic approach with standard questionnaires and Likert scales is important. When face
validity is established one must ask if the population that is being sampled to establish validity is representative of the individuals most likely to use the simulator. It is therefore important to have both expert and novice groups evaluated. The novice group is the group that will actually be using the simulator for acquisition of skill in the future and as such, without their buy in that the simulator is a realistic representation of “real life” experience, further validation is likely not worthwhile. The expert group is critical, as they are the ones who have the clinical experience to be able to determine whether in fact the surgical tasks being simulated can relate to real endoscopic sinus and skull base surgery. In addition, the experts will be involved in the development of surgical simulation curriculum and modules. Face validity was demonstrated using the 11 item questionnaire and the results demonstrate that both experts and novices feel that the simulator is a realistic tool to perform selected simulated tasks that would be encountered during endoscopic sinus and skull base surgery. There was no difference in the 11 items between experts and novices. The respondents felt that the quality of the bone drilling was the most realistic and the properties of the soft tissue and cartilage although realistic, rated slightly lower than some of the other domains that were evaluated. The ratings for realism in our sinus/skull base simulator are comparable to similar studies done by other authors. Bajka and colleagues looked at realism of a hysteroscopic surgical simulator and evaluated face validity in 62 Gynecologic surgeons. All of their respondents found the simulator to be “somewhat realistic”, “realistic” or “absolutely realistic” when evaluating their overall impression. In comparison, in our study we found that 96% of the respondents had this level of realism attributed to the overall impression. Comparison to other simulators that have reported face validation is difficult due to the lack of uniformity in the face validity assessment tools. Despite this, comparisons to others in the literature would suggest that our face validity is as robust as others that have been reported.
The next two domains which evaluated validity, both concurrent validity and construct validity were established using the OSATS tools for assessment based on video evaluation of both the novice and experts performing the surgical tasks on the simulator. Concurrent validity is determined by comparing performance on a new tool/device/simulator with performance on a well established tool/device/simulator of the same construct. In this study we showed that there was no difference in performance between surgeons on the simulator and on the cadaver in 8 of 10 items that were evaluated on the TSS and GRS. The two items that were different between the simulator and the cadaver were the ability to identify the sphenoid face (TSS) and the respect for tissue (GRS). Identification of the sphenoid face proved to be more difficult in the cadaver likely due to the inherent physical properties in a fixated cadaver specimen when compared to the simulator which had more room for identification of the sphenoid. The decrease in respect for tissue in the simulator is difficult to explain but may have resulted from the inherent understanding that the subjects were going to be reusing the cadaveric heads for other components of their dissection, however, the simulated heads were not going to be reused and as such they may have been more aggressive with the tissues. Ritter et al. showed concurrent validity between an established simulator program (fundamentals of laparoscopic surgery (FLS)) and a newly introduced simulator (ProMIS Augmented Reality (AR) simulator). Sixty subjects with varied levels of laparoscopic surgical experience were recruited and performance on the FLS was compared with performance on the AR simulator. Performance metrics included: FLS standard score, instrument path length and instrument smoothness. A strong correlation was found between all three groups of surgeons (novice/intermediate/expert) on both simulators and therefore, they established concurrent validity of the new AR simulator.

The final domain of validity that was demonstrated was construct validity. In order to show construct validity, an assessment tool must be able to differentiate between different levels of
experience. The sinus/skull base simulator was able to show a significant difference in performance based on video review of surgeon performance in 11 of 13 items on the modified OSATS assessment tool for endoscopic sinus surgery. The expert surgeons defined as surgeons who have performed greater than 100 sinus operations had higher scores than the novice surgeons in most domains on the TSS and GRS. Items 4 and 5 on the TSS, although not statistically significant displayed a trend of higher scores in the expert surgeon group.

In a recent study by Fried et al., construct validity was demonstrated on a virtual reality sinus surgery simulator (ES3). Three groups of subjects, that included medical students, residents and expert surgeons were evaluated on three modes of the ES3 that progressed in difficulty. They were able to show differences in performance in the novice and intermediate modes on the ES3, however, there was no difference in the advanced mode. They attributed this to a “learning effect” as one would gain skills in the ES3 as they travelled from one level of difficulty to the next.

In addition to the validity of an assessment tool, the reliability must also be demonstrated. In a systematic review of simulator validation literature, it has been shown that less than half (45%) of studies have reported on reliability. Inter-rater reliability for the modified video-assessment tools on the sinus/skull base simulator was considered substantial to outstanding according to criteria laid out by Landis and Koch. These results further support the reliability scores that were reported by Laeeq and colleagues when the video based assessment tool for ESS was developed. They reported inter-rater reliability scores that were considered substantial.

Hand motion analysis (HMA) and Instrument tracking analysis (ITA) are two other methods of objective analysis of surgical performance. Their role as an assessment tool in endoscopic sinus surgery has yet to be defined. We analyzed surgical instrument movement in both hands. This
included the endoscope, which was primarily manipulated with the left hand, the handheld drill and image guided pointer, which were manipulated with the right hand. The analysis of instruments allowed for the generation of 5 metrics that were mentioned previously. They include time, tool separation, path length, motion smoothness and compactness. To the best of our knowledge, two of these metrics (tool separation & compactness) have not previously been reported on in the arena of surgical simulation. Other studies have shown that motion analysis can be used successfully as an assessment tool for minimally invasive surgery skills. In our study we were not able to discriminate between different levels of experience using the kinematic factors that have been previously described by Cotin et al. or in the two new novel instrument kinematics that we have reported on. The five kinematic parameters that have been previously described include time, path length, motion smoothness, depth response and response orientation. These five factors were then combined to produce and aggregate score that was then shown to have construct validity. It remains unclear which of the specific instrument analysis parameters can be used to differentiate subjects based on levels of experience. The other relevant topic that merits discussion is that each surgical task will produce its own tool/instrument kinematic measurement. Therefore, it is not possible to transfer the validity of one of the surgical kinematics performed during a laparoscopic procedure and state that it will be a valid measurement during a different surgical task, such as ESS. There have also been other studies that have not been able to show construct validity of certain instrument kinematics during laparoscopic simulation.

The two kinematics that we developed: tool separation and compactness during ESS, address the relationship of the tip of the scope to the tip of the surgical instrument (tool separation) and the 3D standard deviation of the surgical tool over time. Both of these kinematics were designed based on the author’s experience with trainees during ESS. It is interesting that the video
analysis allows one to differentiate novice trainees from expert trainees but the instrument kinematics does not. There are a number of possible explanations for this, one of which is the learning curve for motor skill required to complete the task may be quicker than that for the overall task completion itself. It’s possible that a novice surgeon is able to keep their tools separated appropriately during a surgical task, however, they perform multiple unnecessary steps correctly during the task. This will lead to surgical tool metric that appears normal with respect to tool separation but has a poor overall score on the video analysis. As most of the novice surgeons were skewed to have performed between 10-50 or 50-100 cases of ESS, it is possible that the distribution of very novice trainees was underpowered to detect a difference. It’s our feeling that the surgical tool kinematics will be a more valid metric for very novice trainees. This is in keeping with Kopta’s original description of surgical skill acquisition from cognitive to autonomous stages.9 It’s also possible that the kinematics that were analyzed, including path length, motion smoothness, tool separation and compactness are not sensitive enough to pick up differences in experience during ESS despite showing validity in laparoscopic simulators. The paranasal sinus area is a narrow anatomical corridor which prevents large vectors in movement based on a restricted working environment, compared to the abdominal cavity. Based on this alone, the variation in scores may be limited. Ahmidi et al., combined eye motion tracking with electromagnetic tool tracking and found that when both of these metrics were combined during ESS they were able to accurately predict levels of experience.127 This holds promise for future objective assessment of surgical skill during ESS as it would appear that tool kinematics alone may not be sufficient.

Hand motion analysis is another metric used as an objective assessment tool. In a recent systematic review looking at hand motion analysis in laparoscopic training there were a total of 5 systems that were commonly reported on.128 The three most common parameters evaluated
were time, path length and number of hand movements. Construct validity of the HMA systems was the most commonly reported on form of validation (92%). No studies reported on predictive validity and there were a small number of studies that reported on either concurrent validity or face validity (8%). There is a paucity of HMA work in the field of OHNS. Leung and colleagues looked at HMA in a low-fidelity sinus surgery model, however, they were not able to show construct validity for this objective metric. Unfortunately, due to technical issues with the EM system we were not able to get reliable data during the trial. This is one of the ongoing issues with the use of EM systems and includes interference with other electric devices such as the shaver handpiece and drill, as well as “cross-talk” between the wires connected to the EM hand sensors. Improvements to these systems are being investigated within our lab, as well as in the wider image-guidance community.

In summary, there is a possibility that the sinus/skull base simulator could be used as a tool to develop technical skill in endoscopic sinus/skull base surgery. The simulator and video based assessment tool allowed for differentiation of level of expertise (construct validity), showed realism (face validity) and was similar to the existing gold standard for training (concurrent validity). This holds promise for the future of simulation within ESS, as the 3DP process continues to evolve and become more refined, simulator development will continue to advance. Further studies are required to assess and validate instrument tracking metrics and hand motion analysis within the field of endoscopic sinus/skull base surgery.
Chapter 5

5 General Discussion

In summary, these two studies have shown that rapid prototyping, specifically three dimensional printing, allows for creation of accurate and realistic part-task trainers for use in surgical simulation of sinus and skull base surgical maneuvers. The process of three dimensional printing human models is a relatively new concept that requires process validation prior to its implementation and adaptation in surgical training programs. We have shown through qualitative analysis that there are specific printing compounds that have the ability to replicate both bony and soft tissue tactile properties. In addition, the optimal thickness of anatomical membranes and walls were evaluated and defined to aid in the production of models that will be accurate and receptive to conventional endoscopic surgical instrumentation. We also were able to display the accuracy and reproducibility of the printing process with the mean relative difference of the printing process comparing favorably to other forms of RP. Cut off points for membrane thickness and accuracy were established, which is critical for the sinus and skull base region where many of the anatomical structures take on a membrane/wall form. The concept of modulation or printing fidelity was established and to the best of our knowledge is the first time that it has been reported on in the prototyping literature. This allows for an understanding of the minimum distance two objects can be separated and still allow for each to be printed accurately without any fusion between the two anatomical structures. In regions where anatomical detail is precise and critical anatomy is separated by only a few millimetres, a well-defined modulation is critical to realistic model development.
In the second study we were able to validate the sinus/skull base simulator as a tool that can be used to simulate surgical tasks that would be encountered during sinus and skull base surgery. We were able to demonstrate three forms of validity to support this theory. These included face validity or level of realism, concurrent validity which is the comparison to pre-existing established assessment tools, and construct validity, the ability to differentiate levels of experience.

In addition, we defined two new metrics for evaluation of surgical instrument motion during endoscopic procedures. These were compactness and tool separation. Both metrics would appear to be objective assessment metrics that could be used to differentiate surgical aptitude based on experience. However, we were not able to show any difference in performance of novice surgeons compared to expert surgeons in these two metrics. There were no differences seen in other previously described surgical tool metrics; path length and motion smoothness. Thus validity of these objective metrics still requires further evaluation.

The ideal simulator should be safe, reproducible, readily available, portable and generally more cost effective than an animal model or cadaver. Cost-effectiveness was a major factor in our desire to create a sinus/skull base simulator based on 3D printing. Laparoscopic and virtual reality based simulators can have significant cost associated with their acquisition and development. Cost estimates range from $5000 for simple laparoscopic simulators to upwards of $200,000 for more complex virtual reality simulators and anesthesia mannequin based simulators. The cost to produce one of our sinus/skull base models for use in simulation is approximately $300 CDN dollars. This includes the labor and raw material cost to assemble and prepare the model for use. The ability to create realistic high-fidelity models that replicate native
anatomy in a cost effective manner is one of the most attractive features of 3DP. In addition, the digital construct can be altered at no cost, to create models that have different anatomical variants and potential disease states such as tumors or congenital deformities. This allows for the ability to create a wide variety of simulated environments for trainees and novice surgeons to practice on. Other advantages of the sinus/skull base simulator that we developed from 3DP are portability and availability. The drawbacks of cadaveric dissection include accessibility, as the number of specimens is limited by supply and proximity to anatomic labs. The storage and shelf-life of cadaveric specimens is also limited to a pre-defined time period. These issues do not affect our model as it is an inert substance and can be stored indefinitely and used in a number of different clinical settings such as the lab or clinic, without the need for a wet lab for disposal.

One of the drawbacks of both cadaveric and 3DP models for simulation is that if the tasks that are being simulated are ablative and require the removal of tissue from the model then it becomes a single use model. Virtual reality based simulators in contrast, are re-usable for ablative tasks such as ethmoidectomy or drilling of the skull base and over time will recuperate some of the up front financial burden associated with purchase of a virtual reality based simulator. However, there are a number of tasks that can be simulated in the 3DP based model that do not involve destruction of the simulator and thus allow for it to be reused time and time again. Appropriate design of a model to allow for replaceable sections which undergo ablation (e.g., surface of sinuses), while reusing the surrounding exterior (e.g., maxillofacial anatomy) could also help to reduce the costs of a 3DP model.
Chapter 6

6 Conclusions

This study has shown the feasibility of creating a sinus/skull base simulator based on three-dimensional printing. We were able to define the printing compounds that would most closely resemble native sinus and skull base tissue. The accuracy, reproducibility and fidelity of the 3DP process was defined to allow for confidence in the design and creation of anatomic models and to understand the limitations in the fidelity of the 3DP process. Subsequent to the process validation of 3DP, we validated the sinus/skull base simulator based on objective video analysis of novice and expert surgeons performing pre-defined tasks on both cadaveric models and our 3DP model. We were able to define two novel surgical tool kinematics, tool compactness and tool separation, however, we did not show any difference in these metrics or other surgical tool metrics between novice and expert surgeons.

Simulation based training has become a critical component to all surgical training programs and is mandated by certain regulatory educational/licensing bodies such as the accreditation council for graduate medical education (ACGME). The challenge facing surgical subspecialties that have smaller programs with niche areas of surgery is to actually develop and validate surgical simulators that are affordable, portable and accessible. 3DP is an exciting new technologic advancement that allows for creation of models that can be used for simulation and evaluation of new technologies and assessment tools, while fulfilling the previously mentioned principles.

The challenge will be to take these novel forms of simulation and incorporate them into our training programs to create consistent objective metrics based training and assessment that will allow for more efficient training of surgical residents in this new era of medical education. With
work hour restrictions already implemented in most academic training programs, the onus is on surgical educators to achieve a maximal yield from every hour spent at the hospital; surgical simulation with the aid of 3DP will help us accomplish this.
Chapter 7

7 Future Directions

Although this study has defined the parameters for successful creation of a sinus/skull base simulator based on 3DP and validated its use as tool for simulation of sinus and skull base surgical maneuvers, there are still questions that remain to be answered. Future studies will need to show whether training on the simulator as a novice surgeon will translate into improved performance in the clinical theatre. This is the gold standard of simulation and is referred to as predictive validity. It has been reported that less than 5% of studies in surgical simulation have actually addressed this facet of validation.\(^9\) The challenge facing studies of predictive validity is the metrics that are used for summative assessment may not necessarily be representative of improved clinical outcomes for patients. Fried and colleagues displayed predictive validity of the ES3 virtual reality sinus simulator. Their metrics included, injection time, dissection time, surgical confidence, instrument manipulation dexterity and navigation errors.\(^2\) Although all of these factors represent objective metrics to compare two groups of trainees, the argument can be made that they have no impact on relevant patient outcomes. Specifically, improvement in sinus related quality of life or in reduction of complication rates with endoscopic sinus surgery. Major complications in endoscopic sinus surgery are very rare events with estimated incidence of less than 0.2%.\(^3\) A study that evaluates the incidence of complication rates following introduction of surgical simulation will be near impossible to design due to the sample size that would be required to show a significant effect. This leads us down a different path where the outcomes we will be trying to achieve will be proficiency in a defined skill, as opposed to reduction in complication rates or improved surgical outcomes. We are moving toward an era of training to
competency and away from era of training by opportunity. Simulation will help us define what competency is and the metrics we will use to define it.

In addition to competency gained with acquisition of technical skill from simulation, there are other opportunities to incorporate the simulator into sinus and skull base training programs. One of these specific areas revolves around critical incident/disaster management. The most feared complication within sinus and skull base surgery is a rupture of the internal carotid artery while operating in the sphenoid sinus. This is a rare catastrophe estimated to occur in 0.3% of endoscopic skull base surgical cases\(^{131}\) and is very difficult to manage even in the most experienced high volume centers. It requires a careful co-ordination of multiple disciplines including the surgical team, anesthetic team, nursing and interventional radiology. The ability to create an environment that can simulate this rare event has been described by Valentine and Wormald, whereby the internal carotid artery is isolated in a sheep and then a sinus model is placed overtop of the vessel to recreate the hemodynamic environment of this devastating complication. They have shown this is a feasible model to train in the technical skill of carotid hemorrhage management.\(^{132,133}\) The challenge with transferring this exceptional model to other institutions is readily apparent. There are issues with cost, accessibility to animal models and logistic and ethical concerns with this type of simulation. The sinus and skull base simulator that we have developed can be modified to include an active vascular system that can be used to recreate carotid injuries and will serve as an important tool for training of skull base teams in management of this significant vascular injury. The skills gained will not only include technical ones but team collaboration, communication and resource mobilization.

In order to help with material selection and realism of the simulator, the evaluation of synthetic material used by the 3DP process could be more objective. This could be achieved with
biomechanical evaluation of the cartilage and bone found in the paranasal sinuses of cadaveric specimens. This has been performed in human and porcine auricular cartilage models. The common hyperelastic constitutive laws (widely used to characterize soft tissue mechanics) have been described for cartilage found in other anatomic locations. This information would allow not only for subjective quantification of realism but objective biomechanical comparisons as well.\textsuperscript{136}

The other potential avenue for future work revolves around efforts at development of novel technologies to aid in endoscopic sinus/skull base surgery. We have previously reported on augmented reality and virtual reality based navigation systems and their utility both during pre-operative planning of skull base approaches and during simple surgical task completion in cadaveric models.\textsuperscript{82,134} We have shown that surgical “no fly” zones and proximity alerts can be helpful during endoscopic skull base surgery\textsuperscript{135} and the technical aspects of these systems can be further refined and evaluated on the simulator. In addition, metrics looking at surgical tool kinematics that were not validated in this study can be further refined and evaluated. As novel technologies such as 3D endoscopy and eye tracking capabilities emerge the need for validation and reliability testing will become paramount. The decision to incorporate these expensive tools into our continually financially stressed health care system will require metrics to justify their purchase. These can be studied and attained with work on sinus/skull base models generated with 3DP.

These advancements will require the continued close collaboration between surgeon/clinicians and biomedical engineers in order to take the rapidly evolving technical landscape of simulation and three dimensional printing to new heights.
Chapter 8

8 References


7. Munro MG. Surgical simulation: where have we come from? Where are we now? Where are we going? *Minim Invasive Gynecol.* 2012 May-Jun;19(3):272-83.


