Title: Improving Patient Safety in Orthopedic Trauma Surgical Training

Principal Investigator: Donald D. Anderson, PhD

Team Members: Geb W. Thomas, PhD
J. Lawrence Marsh, MD
Matthew D. Karam, MD
Clarence Kreiter, PhD

Organization: University of Iowa

Inclusive Dates of the Project: 1-Sep-2014 to 31-Aug-2017

Federal Project Officer: Kerm Henriksen, PhD
Senior Advisor, Human Factors and Patient Safety
Center for Quality Improvement and Patient Safety
Agency for Healthcare Research and Quality

This project was supported by grant number R18HS022077 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Grant Number: 1R18HS022077-01A1

Structured Abstract (250-words maximum)

Purpose: The goal of this work was to improve patient safety and outcomes by better understanding orthopedic surgical techniques and providing effective simulation tools. The specific objective of the research was to measure the effectiveness of using a radiation-free augmented reality simulator to train residents in wire navigation, a common task in orthopedics.

Scope: This research aimed to provide rigorous scientific evidence that training with a surgical simulator can improve an orthopedic resident's operating room (OR) performance of a common task. This involved work with the simulator as well as performance assessment in the OR.

Methods: Multi-institutional studies involving first-year residents as subjects were conducted to determine how deliberate practice using a radiation-free simulator improves wire navigation performance on a surgical simulation involving actual fluoroscopy. Follow-on studies aimed to demonstrate that skill assessments with the simulator distinguish junior from senior residents. Finally, OR performance measures of wire navigation skills were piloted, established, and validated.

Results: The undertaken studies link training and assessment in the skills lab to performance in the OR. They showed that training on a simulator can improve performance in a mock OR. They also showed that objective performance measures can be reliably measured in the OR. This paves the way for developing, testing, and improving simulators within orthopedic surgery. Increased resident skill in wire navigation gained outside of the OR will spare patients undue additional risk, lead to faster procedures, yield better surgical results, and ultimately provide better patient outcomes.

Key Words: simulation, wire navigation, orthopedic training

Purpose

The **goal** of the funded work was to improve patient safety and outcomes by better understanding orthopedic surgical techniques and providing effective simulation tools. The **specific objective** of the research was to measure the effectiveness of using a radiation-free augmented reality simulator to train residents in wire navigation, a common task in orthopedics. Our **central hypothesis** was that the simulator reproduces critical elements of actual wire navigation. This hypothesis was tested and our objective accomplished by pursuing the following aims.

Specific Aim 1: Demonstrate how deliberate practice improves wire navigation performance.

Multi-institutional studies were undertaken to measure the learning curve for acquiring wire navigation skill on the radiation-free simulator. Prior to progressing to the OR, studies were conducted to show that simulator performance improves performance on a surgical simulation involving actual fluoroscopy.

Specific Aim 2: Show that simulator performance correlates with performance in the OR.

Multi-institutional studies were conducted to demonstrate that skill assessments with the simulator distinguish junior orthopedic residents from senior residents. Then, we piloted, established, and validated OR performance measures of wire navigation skills. Finally, studies were conducted to evaluate how these new OR performance measures correlate with experience.

Scope

Simulation is vital to training and assessment in many surgical disciplines but not yet in orthopedics. Orthopedic surgical skills are largely acquired through apprenticeship in the OR. This is an expensive and unforgiving training model. The paradigm struggles to ensure patient safety. It struggles to provide trainees the variety of experiences they need. The long-term goal of our work is to improve patient safety and outcomes by better understanding orthopedic surgical techniques and providing effective simulation tools. Critical elements needed for progress in this area are the demonstration of simulators that predict OR performance, evidence that they can improve OR performance, and data to support their use as a platform to develop consensus resident performance criteria. The research we have performed is significant, because it enables instructors to confidently quantify resident skill level and efficiently identify and target training to address any skill deficiencies. It also enables standardization of qualifying skill levels across institutions. The development of a strong standard for simulation in the field presents a benchmark for other developers, which will facilitate and strengthen the development and dissemination of simulators within orthopedics. Furthermore, tying performance to the OR represents a critical step forward in the field of medical simulation.

The research conducted took full advantage of exciting and unique new technologies developed in our laboratory to address the important challenge of surgical simulation for training of orthopedic trauma surgical skills. It also produced skills assessment tools, an area in which innovation was likewise desperately needed. Currently, orthopedic accrediting bodies, and even orthopedic educators who work with trainees on a daily basis, have no objective metrics to assess and guide learners and to ensure that learner competence is being or has been attained. Furthermore, inexperienced surgeons and trainees cannot frame their own progress against benchmarks or identify areas in which they need practice. Our work has demonstrated that a relatively common task, wire navigation, can be simulated in a laboratory environment and that resident performance can be objectively measured. We have worked to demonstrate the generalizability and validity of our results. Laboratory-based models of orthopedic tasks and operations, combined with objective assessments of performance such as those presented here, hold great promise as tools to improve resident training in surgical skills and to make valuable OR experiences more safe and effective.

This research built upon our initial experience using a custom-developed wire navigation simulator at a single institution, with a limited number of orthopedic residents available to participate. Our research aims could not have been satisfactorily achieved without the experimental power and generalizability afforded by pooling research participants from a consortium of residency programs. In order to manage research costs, without great sacrifice to the variety of participants that could be recruited, we decided to emphasize regional medical centers. To that end, co-investigators Drs. Marsh and Karam invited the participation of this project within their recently established Midwest Orthopedic Surgical Skills (MOSS) Consortium.

The nine-school MOSS consortium of orthopedic residency programs was formed to respond to a mandate from the Orthopaedic Residency Review Committee and American Board of Orthopaedic Surgery (ABOS), requiring a surgical skills training program for all Post-Graduate Year 1 (PGY1) orthopedic residents. The purpose of the consortium is to develop and refine approaches for resident surgical skills training and assessment, with an emphasis on orthopedic trauma surgery. All the participating institutions host 5-year residency programs. We invited programs from this consortium to participate in the proposed experiments. In the end, four programs participated: University of Minnesota, Mayo Clinic, University of Nebraska, and University of Iowa. Given difficulties in traveling around the Midwest and in getting local programs to participate fully in our work, we turned our attention near the end of our funding period to participation in regularly scheduled fracture courses offered by the Orthopaedic Trauma Association (OTA). The OTA hosts this course twice a year, with approximately 100 first and second year orthopedic residents participating. An expert faculty from across the U.S. is assembled to teach the course, affording us additional opportunities to evaluate expert performance.

Methods

Demonstration and dissemination of an orthopedic trauma simulator

Over the period during which AHRQ funded this research, we more fully developed an orthopedic trauma surgical simulation platform that replicates the look and feel of navigating a wire through bone utilizing fluoroscopic guidance (Figure 1). Two differentiating features of the platform are (1) camera-based tracking of a surgical wire to replace fluoroscopic radiation exposure and (2) a foam bone surrogate to replicate the feel of drilling through actual bone. We initially developed this platform to

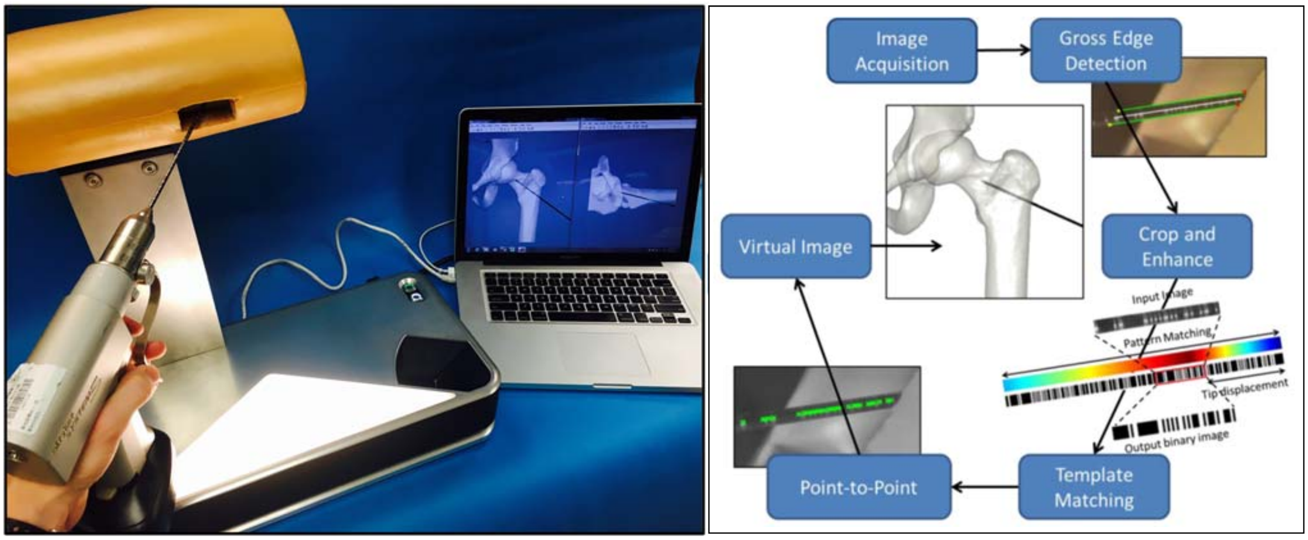

Figure 1. The orthopedic surgical simulation platform (left) relies upon camera-based tracking of a laser-etched stainless steel K-wire (right) identical to that used in surgery. In the case of the hip wire navigation simulation shown, a foam bone surrogate is housed within a soft tissue mimicking sleeve to obscure the trainee's view of the object.

simulate the navigation of a wire in the treatment of intertrochanteric hip fractures. A small company, which we founded, has licensed the technology from the University of Iowa Research Foundation in order to inexpensively produce the simulator for other educational institutions. This hip fracture wire navigation simulator is currently being tested at orthopedic residency programs across the Midwest, and the technology has been substantially improved along the way.

The current device is more precise and less expensive than earlier prototypes. Figure 2 shows the first field test of the current version of the simulator in August 2016. A resident is holding a battery-powered surgical drill with a specially prepared surgical wire. He is driving the wire into a plastic bone covered by a rubber soft tissue envelope. The bone is secured to the vertical mast of the simulator. Both the resident and the graduate student operator are looking at a rendering of the simulated fluoroscope views (Figure 3) on the laptop computer. The resident uses these views to guide his decisions about how to control the drill as he or she attempts to achieve the correct position of the wire along the central axis of the femoral neck, with the tip of the wire just inside the curve of the femoral head.

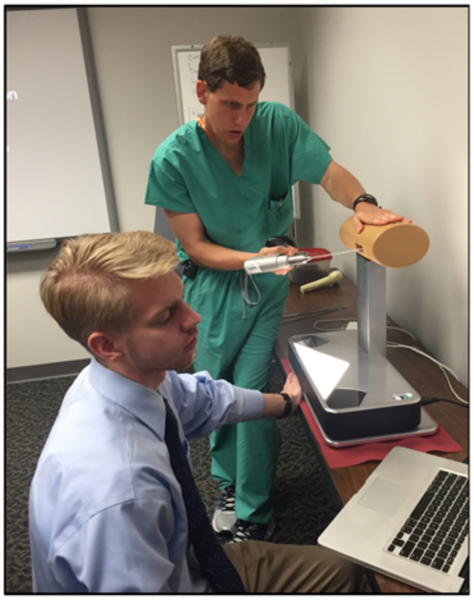

Figure 2. A resident using the wire navigation simulator.

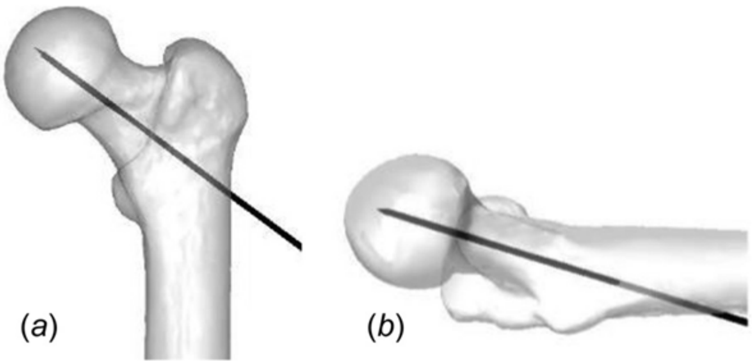

Figure 3. Simulated fluoroscope images showing the (a) AP view and (b) lateral view of the proximal femur produced by the wire navigation simulator. The position of the wire mirrors the physical position of the wire relative to the plastic bone.

The simulator is portable. The mast upon which the surrogate bone is mounted can be removed, and the device can be packed inside a hardshell case (Figure 4). The replaceable plastic bones and soft tissue envelope are manufactured by Sawbones, Vashon Island, WA. The simulator itself may be purchased from Iowa Simulation Solutions, LLC, which has licensed the patented technology from the University of Iowa, in order to make the simulator more broadly available to the orthopedics community.


Figure 4. Simulator packed in hardshell case for travel.

The simulator uses an integrated camera, an on-board Raspberry Pi miniature computer, a personal computer, and image processing methods to track a laser etching-addressed guide wire relative to a fixed surrogate bone model. The camera and an arrangement of mirrors (Figure 5) provide two views of the wire as it enters the bone. The image is passed to a laptop computer, which processes the image (Figure 1, right-side graphic). First, it divides the raw image into two sub-images, each viewing the wire from a different position, one from the lower left, the other from the lower right. It then uses Matlab image processing functions to find the edges of the wire in each image. Given the calibrated position of the camera views and the known position of the bone on the mast, this analysis yields the 3D vector of the wire relative to the bone. A black stripe pattern similar to a barcode is laser etched on the wire. The visible portion of the wire is compared to a reference image of the wire to determine the approximate position of visible portion relative to the wire reference. A detailed analysis of the stripes then reveals the exact position of the wire tip relative to the wire vector. This algorithm provides a complete description of the wire position relative to the bone, accurate to approximately 1 mm and within 2° (details below).

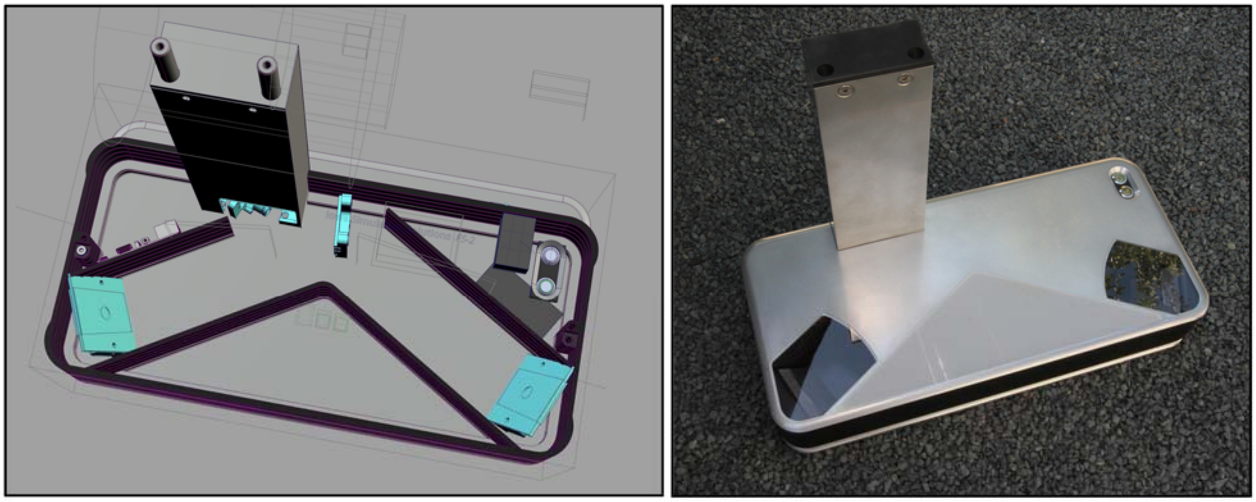
Figure 5. A camera hidden under the mast views two optical paths to tilted mirrors under the glass at each corner. The path to the lower left corner is enabled by the mirror at the center back that redirects half of the camera view. Together, the mirrors allow the single camera to simultaneously achieve two views of the wire.

In subsequent work, we explored the potential for our simulation platform to accommodate other surgical scenarios, demonstrating along the way that a variety of tasks may be readily simulated. We also quantified the system accuracy, which is an essential component for defining new surgical scenarios. The line-of-sight constraints of an optical system effectively restrict the space through which a surgical wire can be tracked. This work envelope is a frustum extending about 4 inches from the surgical window. The small end of the frustum is approximately the size of the window, and the large end is roughly twice as long and high. A second defining characteristic of the system is the precision with which a wire can be tracked. If a surgery to be simulated demands greater precision than the simulator can provide, it will not be able to provide a satisfying experience for the learner. Finally, as the wire travels through the simulated bone, the user feels resistance to forward and rotational movement.

Whether or not the simulated bone material must have a shape that is realistic is not entirely clear: at some point the resistance to forward movement is sufficiently large that it is unlikely that the learner can feel the harder material at the back edge of the bone. If the back edge cannot be felt while drilling, it may be sufficient to simply provide enough material to provide the necessary masking resistance at a given drill depth.

Any surgery that falls within the constraints of work envelope, precision, and bony material representation can be readily simulated. Potential candidates include sacroiliac screw fixation, distal radius fracture fixation, the fixing of pediatric elbow fractures, the placing of pedicle screws for spinal fusion, fixing proximal humerus fractures, and tibial plateau fractures. As an example, we modified the simulator to represent the placement of a sacroiliac screw. In this surgery, a long screw must be placed into the pelvis crossing the sacroiliac joint. The spinal cord and other central nervous structures nearby must be avoided. In the case of a sacroiliac screw fixation module, the main views used are referred to as pelvic inlet and outlet views (Figure 6). We have also developed a preliminary demonstration that uses the device for pediatric distal elbow fractures.

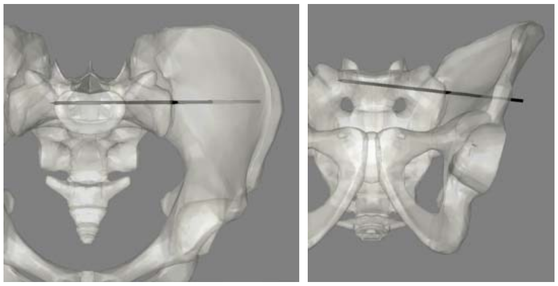
Figure 6. Pelvic inlet (left) and outlet (right) views of the model for sacroiliac fixation.

For effective training capability, the wire navigation simulator must accurately map motions of the learner to the virtual fluoroscope images. Because it is a vision-based system, it is sensitive to ambient lighting conditions. To measure the effect of ambient light on accuracy and to quantify the system accuracy in general, a testing bone was created with known wire positions and trajectories (Figure 7). These wire positions and trajectories were obtained by placing the wire in bone and acquiring a laser scan of the setup. The accuracy of the simulator was then tested at a variety of different locations by placing the wire in the target bone and comparing the known tip position and wire trajectory to the calculated pose. The time taken for the software to perform the pose calculations was also recorded for each location.

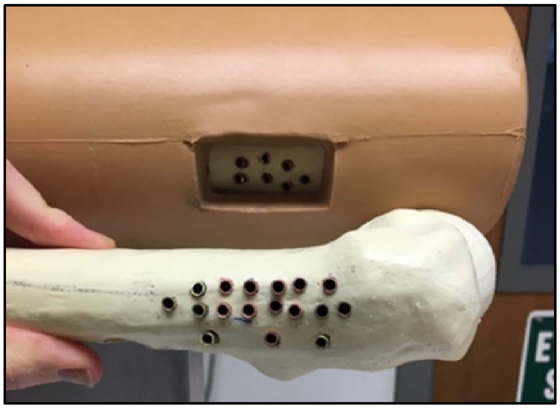
Figure 7. The surrogate bone is shown here, with 17 small diameter brass tubes inserted to provide reproducible wire placement.

At each testing location, 290 accuracy assessments were performed (29 slots, 10 tests per slot). In these readings, the wire tip accuracy varied between 0.25 mm and 4.85 mm of error (Table 1). On average the wire tip error was 1.7 mm. The wire angle error varied between 0.04° and 4.3°. The average angle error was 1.31°. Across all tests, the average time to compute the pose was 1.05 seconds, well within the range of a typical fluoroscopic image acquisition in surgery (~25 seconds). This work was completed early in our study. Since then, we have made several improvements to the system's lighting and focus that has slightly enhanced the system accuracy and considerably improved its robustness.

Table 1. Simulator Accuracy

Location	Tip Error (mm)	Angle Error (°)	Computation Time (sec)
Desktop	1.53	1.15	1.12
Dry Lab	1.59	1.31	1.03
Conference Room	1.61	1.12	1.01
Wet Lab	1.97	1.18	1.08
Library	1.81	1.31	.99
Average	1.7	1.31	1.05

This work demonstrates the versatility of the extensible wire navigation simulation platform. As programs begin to expand their use of simulation from laboratory training for first-year residents to assessments of competency for more senior residents, a wider array of tools will be needed. Being able to incorporate more difficult surgeries, such as pelvic or spine procedures, will provide program directors and governing bodies the tools they need to properly assess surgeons as they move toward certification. Furthermore, with these more difficult procedures, there is clearly a greater need for training outside of the OR. Training on a simulated pelvic or spine model will allow surgeons to better understand the mistakes that can be made in those procedures before they enter the operating room, leading to improved patient safety.

The American Board of Orthopaedic Surgeons, the principal accrediting body for the orthopedics community, invited us to demonstrate the device in August 2016, as they begin to formally consider adopting such a simulator for assessing resident surgical skill. We then visited with them on-site during their annual certification exams in July 2017. They have indicated a willingness to pilot our system in the coming year. We also have a commitment from the Orthopaedic Trauma Association, one of the largest professional associations within the orthopedics community, to partner with them at their bi-annual resident training courses in the coming years. Through these programs, we expect to introduce hundreds of residents to the simulator over the next 5 years, collecting data that will provide unique scientific validation of a training approach that we expect to become the standard within the next decade.

In order for simulation to gain acceptance as a fundamental training tool, particularly if it is to have any formal role in accreditation, it is likely that the community must be assured of the simulation's ability to impact operating room performance. Scientific progress here depends on well-defined, sensitive, objective measures of operating room performance. In our AHRQ-funded work, we evaluated performance on the simulator in both a mock OR and in the real OR (Figure 8). This allowed us to differentiate performance improvements caused by simulator training from those improvements that accrue with experience. Subtle tests with few participants require sensitive performance measures. To maximize yield, we define consistent measures common to both the simulator and the actual OR.

Figure 8. The current orthopedic surgical simulator (left). Performance is measured in a mock OR (center). The purpose of this project is to measure resident performance from information we've collected in the real OR.

Objective and reliable evaluation of operating room performance

There are no widely accepted, objective, and reliable tools for measuring surgical skill in the operating room (OR). Ubiquitous video and imaging technology provide opportunities to develop metrics that meet this need. Hip fracture surgery is a promising area in which to develop these measures, because hip fractures are common, the surgery is used as a milestone for residents, and it demands technical skill. Thus, one study objective was to develop meaningful, objective measures of wire navigation performance in the OR.

We collected videos of OR performance along with relevant fluoroscopic images and general case data, hosting 45 cases with our custom, password-protected web-server software and making them available to residents and staff. As a result of nearly 2 years of regulatory and technical work, University of Iowa orthopedic residents wore a head-mounted, GoPro, point-of-view camera while performing surgery. Every fluoroscopy image collected during these surgeries was saved to the patient's electronic medical record. Residents could immediately upload and view their videos. The fluoroscope images were superimposed on a corner of the video synchronized in time with the video, supporting the surgical narrative (Figure 8, image to right of figure).

In one study we conducted using this approach, resident surgeons at the University of Iowa Hospitals and Clinics wore a head-mounted video camera while performing surgical open reduction and internal fixation using a dynamic hip screw (Figure 9). Performance metrics were drawn from the video as well as from intraoperative fluoroscopy images that were obtained during the surgery and saved for later analysis (Table 2).

Figure 9. Intraoperative picture of a resident operating (left) and camera view during wire navigation (right).

Four metrics were derived from the data. The first metric was an objective, geometric measurement of the wire placement compared with a defined ideal placement. The second was the time to conduct the procedure. The third was the number of fluoroscopic images collected. The fourth metric was the degree of supervisory intervention, an indicator of resident skill and readiness for independent practice. The second metric is unavoidably sensitive to the interaction style of each supervisor. Any OR performance assessment involving residents must account for intervention by the supervising surgeon that influences the progression of the procedure. The degree of intervention of the resident's supervisor was measured by tallying the weight of all interventions, in which each intervention was given a weight of 1 – supervisor instructions to someone other than the operating surgeon (e.g., requesting a fluoroscopic image from the radiology technician), 2 – supervisor instructions to the operating surgeon (e.g., instructions to alter the wire trajectory), or 3 – the supervisor handling an instrument in a capacity exceeding that of an assistant (e.g., physically controlling the wire guide; Figure 10). For trend analysis purposes, the supervisors were grouped into two categories; a senior-level resident or a staff surgeon.

Table 2. Summary of Scoring Metrics and Source

Metric	Unit	Video	Fluoroscopic Image
Tip-apex distance (TAD)	mm		X
Duration of procedure	Minutes	X	X
Fluoroscopic shots	Count	X	X
Pullbacks	Count	X	X
Head cortex breaches	Count	X	X
Switch AP/lateral shot	Count	X	X
Supervision impact score	Formula	X	

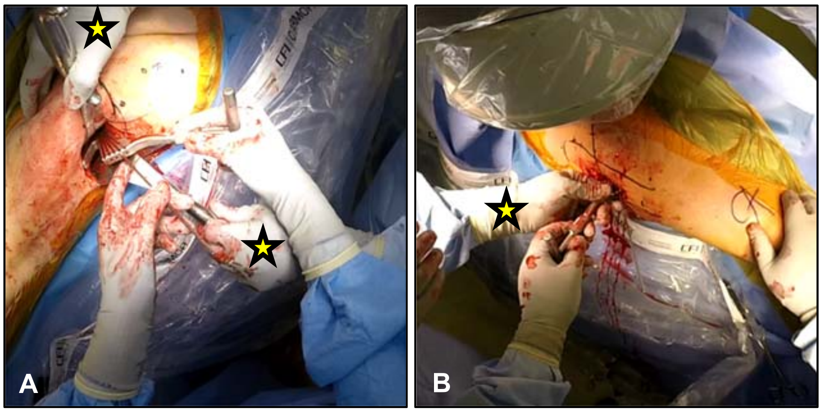

Figure 10. Examples of two supervision intervention behaviors. (A) The supervisor (hands indicated with stars) is handling the Bennett retractors in an assisting manner (elevation and retraction), and no supervision impact is counted. (B) The supervisor (on the left, hand indicated with a star) is taking control of the guide, which constitutes handling an instrument and is tallied with a weight of 3.

To determine reliability of these measurements, four independent non-expert raters performed them for two cases. Raters independently measured the tip-apex distance (TAD), a measure made from fluoroscopic images that reflects the accuracy of the surgical placement of the wire, on all seven cases. Seven surgeries were performed by seven different orthopedic residents. All four raters were biomedical engineering graduate students.

TAD Results

For all graders, the average standard deviation for AP, lateral, and summed TAD measurements were 2.7, 1.9, and 3.7 mm, respectively (Table 3). For the AP measure of case 1, there was one (of 140) extreme outlying measurement (over 3.5 times the average and double the next highest value) that disproportionally skewed the standard deviation. With this single outlier removed, the average AP standard deviation for case 1 was reduced from 6.0 to 2.5 mm and for all AP cases from 2.7 to 2.1 mm. The summed TAD standard deviation for case 1 was reduced from 6.9 to 3.1 mm and the average for all cases, from 3.7 to 3.1 mm.

Table 3. Means and Standard Deviations in TAD Measurements of 7 Cases by 10 Raters

	Mean (standard deviation) (mm)		
Case	**AP**	**Lateral**	**Sum**
1	6.2 (6.0)	4.9 (1.2)	11.1 (6.9)
2	5.2 (1.5)	9.6 (1.8)	14.8 (2.3)
3	8.8 (3.8)	10.4 (3.4)	19.2 (5.6)
4	4.6 (1.3)	3.9 (1.8)	8.5 (2.5)
5	12.0 (1.6)	11.7 (2.6)	23.7 (2.8)
6	8.2 (2.8)	9.2 (0.9)	17.4 (3.3)
7	7.7 (1.9)	9.8 (1.6)	17.5 (2.3)
Average	7.5 (2.7)	8.5 (1.9)	16.0 (3.7)

The inter-rater reliability analysis for all measures together (AP, lateral, and sum) and across the 10 raters produced a Cronbach α of 0.97 and an Intraclass Correlation Coefficient of 0.72 for single measures and 0.96 for average measures. For intra-rater reliability between first and second measures of the same TAD with the two raters who repeated their measurements after 2 weeks, rater 1 had an average standard deviation of 1.39 mm and rater 2 had an average standard deviation of 1.92 mm. Rater 1's Cronbach α and an Intraclass Correlation Coefficient (single measures and average measures) were 0.94, 0.89, and 0.94, respectively. Rater 2's Cronbach α and an Intraclass Correlation Coefficient (single measures and average measures) were 0.88, 0.79, and 0.88, respectively.

Video Metric Results

Task duration (from fluoroscopic timestamps), number of fluoroscopic shots, switches between AP and lateral views, and cortex head breaches were consistent for all raters on both the cases. [Task duration had also been recorded by raters using video timestamps (results not shown); this approach for estimating duration produced discrepancies between 2 and 21 seconds for case 1 and between 1 and 25 seconds for case 2, so it was abandoned.] For both the cases, the number of pullbacks, instructions to others in the OR, instructions, and supervising surgeon impact score varied between raters. The average standard deviation for instructions to others, instructions to the operating surgeon, handling of instruments, and impact score for both cases was 1.64. The standard deviation of the number of pullbacks was 0.5 for case 1 and 1.7 for case 2. The inter-rater reliability analysis for all nine metrics across both cases produced a Cronbach α of 0.996 and an Intraclass Correlation Coefficient of 0.90 for single measures and of 0.99 for average measures. The inter-rater reliability analysis for pullbacks alone for both cases produced a Cronbach α of 0.33 and an Intraclass Correlation Coefficient of 0.95 for single measures and of 0.30 for average measures. The inter-rater reliability analysis for the supervising surgeon impact score alone for both cases produced a Cronbach α of 0.99.

The high values of inter-rater reliability (0.969 Cronbach α) and intra-rater reliability (0.881 and 0.943 Cronbach α) show that the technique for measuring the TAD from intraoperative fluoroscopic images is valid and reproducible. The magnitude of the TAD summation standard deviation (3.1 mm) for 10 raters across seven cases seems acceptable, given the utility of the measure and keeping in mind that the error is relatively small when compared with the criteria of 25 mm typically applied for clinical acceptance.

The video metrics exhibited high reproducibility between independent raters. Both cases had the same values across all raters for four of the nine metrics. The number of pullbacks was less consistent than some of the other measures, with an inter-rater reliability Cronbach α of 0.33. We conclude that the number of pullbacks is unreliable and should not be used as an assessment metric. The lack of instances in which the guide wire breached the cortex of the femoral head limits the study of that parameter, as no occurrences were identified from video or intraoperative images for the cases studied. Both pullbacks and cortex head breaches rely on determining a wire's position, which proved to be difficult—the drill tends to obscure the line of sight in the video images and the fluoroscope images are not continuous, so a head breach or repositioning may not be explicitly captured. Objectively determining whether a wire was removed and reinserted into a previous hole in the cortex or versus starting a new cortex breach is difficult, as the C-arm position may change between images.

To summarize, this study included four raters viewing fluoro images and video collected from a head-mounted camera for seven surgeries. The raters independently rated nine metrics from the data. In total, four of these---the surgery duration, the number of fluoroscope images collected, the number of C-arm position changes, and the degree of surgeon intervention (a weighted sum of three categories of intervention)---were consistent across the four video reviewers and are likely to be useful for performance assessment. The TAD was less reliable than previous reports have suggested, but it is still a valuable metric of surgeon skill. Our study shows that video recording assessment allows non-experts to reliably measure these metrics; they offer an opportunity for objective, consistent assessment of OR performance.

Results

Simulation Training Benefits Mock OR Performance

Thus far, 54 first-year residents have participated in our hip fracture wire navigation simulator experiments, with residents drawn from four Midwestern orthopedic residency programs. The task simulates placing a guide wire during the surgical repair of intertrochanteric femur fractures using a dynamic hip screw. Central to that surgery is the skill of wire navigation, which is generalizable to other common surgical procedures. In each mock OR trial, the participant uses a fluoroscope to navigate a surgical guide wire through an opening in a soft tissue-mimicking rubber exterior and into a plastic femur, drilling from the lateral side, through the neck and into the femoral head. The wire is used to accommodate the later placement of a larger cannulated lag screw.

We assess task performance using several measures that were also used in the live OR study, each addressing resident skill and directly relevant to patient safety. The first measure, the TAD, evaluates the accuracy of guide wire placement within the femoral head. The TAD is measured from antero-posterior (AP) and lateral fluoroscopic images, summing the distance in each image between the tip of the lag screw to a point on the femoral head. Screws placed with a TAD greater than 25 mm have a higher risk of mechanical failure. The second performance measure is the task duration. Shorter durations demand economy of motion and lower the risk of infection. The third measure is the number of fluoroscopic images used in placing the wire. Fewer images require a higher degree of planning and spatial reasoning from the resident and expose the resident and the patient to less radiation.

Through our experiments, we have determined that our simulator training improves performance in the mock OR. Forty-two first-year orthopedic surgery residents participated in this portion of our studies. Participants included residents from the University of Iowa, University of Minnesota, the Mayo Clinic, and the University of Nebraska---all members in the MOSS Consortium. Residents were split into three different cohorts that each received varying levels of training.

The 23 residents in Cohort 1 received traditional training followed by a performance assessment. The traditional training included a didactic PowerPoint presentation on placing a guide wire in the treatment of an intertrochanteric hip fracture, then a video from the ABOS module on Fluoroscopic Knowledge and Skills demonstrating the proper technique for placing the guide wire. The performance assessment measured their ability to place the wire through the femoral neck and into the femoral head in a mock OR.

The 13 residents in Cohort 2 received the same traditional training, then simulator training, then a performance assessment. Simulator training included 30 minutes to practice the task using the wire navigation simulator. The simulator provided real-time feedback to the residents on their wire position relative to the ideal wire position while they practiced navigating the wire.

The six residents in Cohort 3 received the traditional training, then proficiency-based training, followed by a performance assessment. The proficiency-based training began with a computerized task identifying the correct wire entry and end point on a series of fluoroscopic images, followed by guided simulator practice breaking down the task into subcomponents, including identifying the proper starting point and wire trajectory in both AP and lateral images. Each element was rehearsed until a predefined level of proficiency was attained.

Table 4. Mock OR results. The p values indicated are tested against Cohort 1.

Cohort	TAD (mm)	P	Time (sec)	P	Images	P
1	23.6 ± 7.3		242 ± 124		21 ± 11.9	
2	17.1 ± 5.3	0.005	226 ± 101	0.69	17 ± 5.9	0.15
3	12.6 ±2.8	< 0.001	357 ± 104	0.06	27 ± 11.4	0.37

Both Cohorts 2 and 3 demonstrated significant improvement in their TAD compared to Cohort 1 (Table 4), indicating that training improved performance. Cohort 3 attained the best average TAD, roughly half that of Cohort 1, but at the cost of more fluoroscopic images and procedure time, suggesting that the training to proficiency changed Cohort 3's speed-accuracy trade off. These results indicate the effect that the different styles of training had on each group. In Cohort 2, simulator practice repeated the same procedure several times, possibly leading to a quick, decisive approach. Cohort 3 was taught to approach the task in a more analytical and algorithmic fashion. Possibly this cohort needed more time and more images in the assessment because they were being more analytical in their approach, yielding a higher mean TAD, arguably the most clinically relevant metric for surgical success. Perhaps with even more time to practice, Cohort 3 would be able to further streamline technique to reduce the time and fluoroscopic images while still maintaining low TAD values.

This past year in an exploratory experiment, we recruited 22 residents from participants at the 2017 Orthopaedic Trauma Association Spring Comprehensive Fracture Course for a study to determine to what extent residents develop wire navigation skills while participating in the course. After familiarization to the simulator and the task, followed by a few minutes of practice to become acquainted with the simulator, each resident completed a pre-course assessment of their wire navigation skills using our simulator. The residents then completed two fracture modules in the Orthopaedic Trauma Association course felt to be most relevant to hip wire navigation: the geriatric and pediatric fracture modules. In general, the course has residents practice drilling various implants into bare Sawbones, with guidance and instruction from course faculty members. Seventeen of the participant residents then completed a post-course assessment on the simulator identical to the pre-course assessment. Resident performance was assessed based on wire position, total time, and the number of images requested. There was no appreciable change in performance between the two trials, which suggests that course modules demonstrating surgical procedures on bare Sawbones do not adequately encompass the complexities associated with placing a wire under fluoroscopic guidance.

We are moving quickly to integrate our simulator and training into two separate modules at the Orthopaedic Trauma Association's biannual fracture course. One module focuses on geriatric fractures and incorporates the intertrochanteric hip fracture scenario we have been studying. The other module focuses on pediatric fractures and incorporates a pediatric elbow (supracondylar humeral) fracture scenario very similar to the hip fracture. Additional challenges presented in the pediatric elbow fracture scenario include the precise placement of multiple wires to achieve a mechanically favorable fixation.

OR measures discriminate performance in agreement with experience level

Although our previous OR measures had a high reliability, we may need even more sensitivity if we hope to see the effect of training within the complex environment of a real OR, complete with experimental confounding factors like patient variability and the presence of surgical assistants and supervisors. Building on our studies in the operating room that established objective and reliable performance measures, we undertook an evaluation of the correlations between these measures and a resident's experience level. Data were collected for 18 hip fracture cases. Three were subsequently excluded from analysis: one because it involved a very unstable fracture that confounded wire placement, another because the supervising surgeon was emergently called away during the wire navigation, and a third because the supervising surgeon was using a new implant for the first time, which substantially increased the duration of the surgery and added discussion time related to the unfamiliar implant. One staff surgeon and 13 different residents functioned as the operating surgeons (one resident completed two procedures, both as a third-year resident). Three residents were women, 10 were men, and the staff surgeon was a man. The number of weeks into residency and the number of comparable hip fracture cases previously logged by each surgeon were used as indicators of surgical experience. The number of cases came from the resident's ACGME case log. The staff surgeon's number of logged cases (24) was taken from a billing log felt to most appropriately represent his considerable prior case experience.

The 15 different hip fracture repair surgeries were recorded using a head-mounted point-of-view camera. Intraoperative fluoroscopic images were also saved. The following performance metrics were analyzed: duration of wire navigation, number of fluoroscopic images collected, degree of intervention by the surgeon's supervisor (see next paragraph for details), and the TAD. Two orthopedic traumatologists additionally graded surgical performance in each video independently using an Objective Structured Assessment of Technical Skill (OSATS) global rating score.

A composite score was computed by summing the average standardized values of the four performance metrics using SAS (version 9, SAS Institute, Cary, NC). For the four cases in which a TAD or supervision intervention score was not available, the composite score was computed by summing the average of the three remaining standardized performance values. Correlations between the composite score and experience were calculated using the same approach as for the individual performance metrics.

Table 5. Correlation coefficients relating performance metrics to surgeon experience. Significant correlations (p<0.05) are bolded.

	Experience Metrics			
Performance Metrics	Weeks in Residency		Cases Logged	
	Correlation Coefficient	p-value	Correlation Coefficient	p-value
Duration (n=15)	**-0.661**	**0.007**	**-0.587**	**0.021**
Fluoroscopic Images (n=15)	-0.340	0.216	-0.268	0.335
Supervision Intervention (n=14)	-0.485	0.079	-0.346	0.226
TAD (n=12)	-0.201	0.531	**-0.669**	**0.017**
Composite Performance Metric (n=15)	**-0.549**	**0.034**	**-0.656**	**0.008**
OSATS (n=13)	0.044	0.886	0.092	0.765

Table 5 presents the correlation coefficients relating performance to experience metrics. The number of previous cases logged for residents ranged from one to 13, with an average of 5.7 (4.1) cases. Wire navigation duration was significantly correlated with both weeks into residency -0.66 (p<0.01) and prior cases logged -0.59 (p=0.02) (Figure 11, left graph). The number of fluoroscopic images and the supervision intervention score did not correlate with either experience metric. TAD was significantly correlated with cases logged -0.67 (p =0.02; Figure 11, right graph) but not weeks into residency. A Mann-Whitney U analysis indicated that the wire navigation duration for the novice group was significantly higher than for the experts (p=0.05). There was no significant difference in the TAD between the two experience groupings. The composite performance metric significantly correlated to both weeks into residency -0.55 (p=0.03) and cases logged -0.66 (p=0.01). The stronger correlation seen between performance and the number of previous cases logged, rather than the point in residency, is consistent with previous findings that surgical skill is not directly related to duration of surgical practice but to the number of surgeries performed.

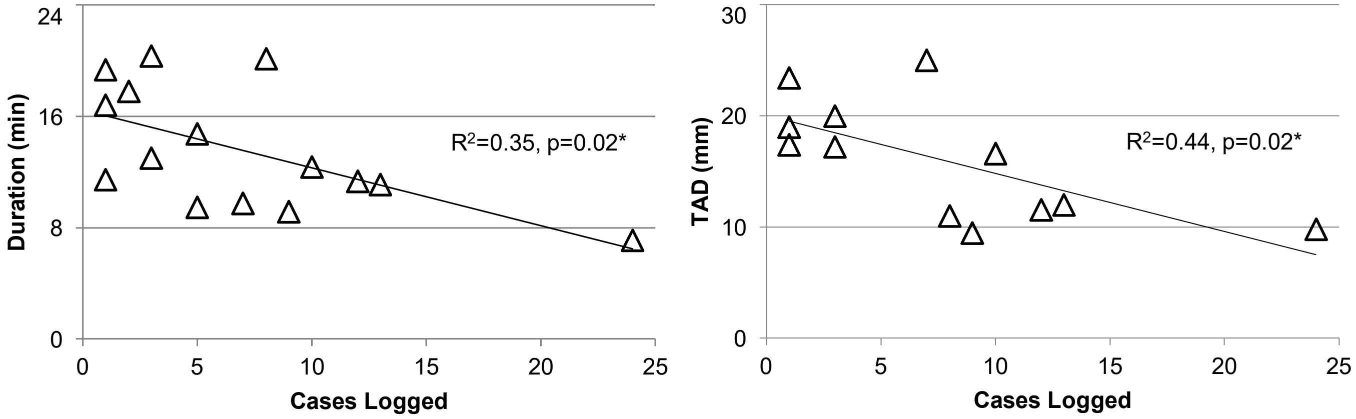

Figure 11. Plots of duration and TAD vs. the number of previous cases logged. Results of linear regression are shown as a general indicator of the relationship.

The inter-rater reliability of the two traumatologists' OSATS total scores was 0.71. There was no significant correlation between the OSATS total score and experience metrics (correlation coefficients: weeks into residency 0.43, p=0.11 and cases logged 0.43, p=0.11). Total OSATS score significantly correlated with duration -0.52 (p=0.05) and number of fluoroscopic images -0.83 (p<0.001) but correlated neither with supervision intervention score -0.27 (p=0.36) nor TAD -0.30 (p=0.34).

The supervision intervention score did not significantly correlate with any of the performance metrics but negatively correlated with weeks into residency -0.49 (p=0.08). This negative trend supports the hypothesis that, as surgeon experience increases, the supervision intervention score decreases, most likely because the resident is more prepared for independent practice. Interestingly, although an increase in supervision intervention score did significantly predict increased duration, there was no corresponding improvement in the TAD.

The results from this study indicate that two individual metrics of hip fracture wire navigation performance, duration and TAD, significantly differentiate surgical experience. A composite score incorporating multiple performance metrics also provided strong correlations with surgical experience. The methods presented have the potential to provide truly objective assessment of technical performance in the OR, a critical step toward ensuring surgeon competency. Task duration had the strongest correlation with weeks into residency (p<0.01), a finding consistent with previous studies showing that greater surgical experience is associated with shorter operations. TAD correlated negatively with the number of previous similar cases logged by the surgeon (p=0.02), agreeing with another study in which experienced surgeons obtained better wire placement than novices.

These preliminary findings suggest that procedure duration and TAD may be useful as objective, quantitative measures of OR performance. They are not completely satisfying, however, because duration combines many potential causes of delay and ascribes these delays to surgical skill. TAD emphasizes only the location of the tip of the surgical wire and neglects the wire's trajectory through the femoral neck. Resolving these challenges may lead us to a richer understanding of what constitutes surgical skill and how to best measure it.

Toward a better measure of surgical wire placement

The inadequacy of the TAD as a measure of wire placement does not only arise from its inability to measure the trajectory of the wire through bone. The inter-observer variability of TAD measurements is as high as 10%. Some of this variability is a result of uncertainty in the position of the fluoroscope and some is the result of the subjective elements of the protocol used to locate the geometric reference features on the fluoroscopic images.

We have developed a new measure, called the trajectory accuracy metric (TAM). The TAM leverages graphical image processing and solid modeling capabilities to express an ideal wire trajectory relative to a specific bony anatomy. The TAM is then defined as the average distance between the ideal and actual wires within the bone (Figure 12). This measure penalizes wire placements that have very different starting positions or do not go through the center of the femoral neck, both of which are desirable clinical goals and indicative of a surgeon's skill in navigating the wire.

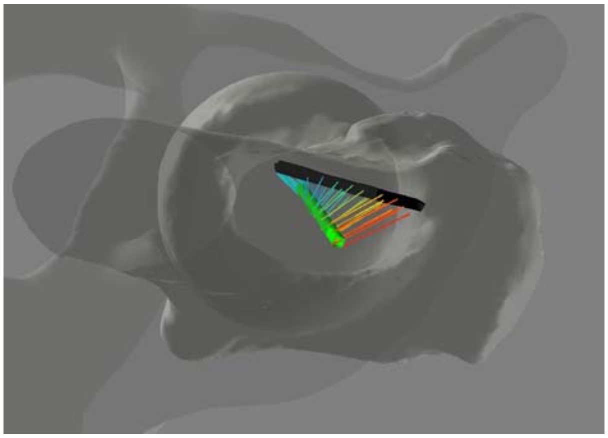

Figure 12. The trajectory accuracy measure (TAM).

To automatically obtain the TAD and TAM from fluoroscopic images, an important advancement for using these methods eventually in the OR, it is necessary to identify the pose of 3D objects as they appear in 2D projection. This requires the use of a pose identification algorithm that relies upon established 3D to 2D model alignment methods. Fortunately, our group has developed substantial capabilities in this area in support of intraoperative surgical navigation efforts. To this point in our surgical skills simulation work, we have only used this measure with respect to our mock OR experiments, in which the geometry of the femur is well known, simplifying the challenge of defining the ideal wire position (Figure 13). However, as we look forward to application in the actual OR, it will be necessary to refine the image processing techniques in order to define an ideal wire position *in situ*, leveraging our experience with similar modeling techniques.

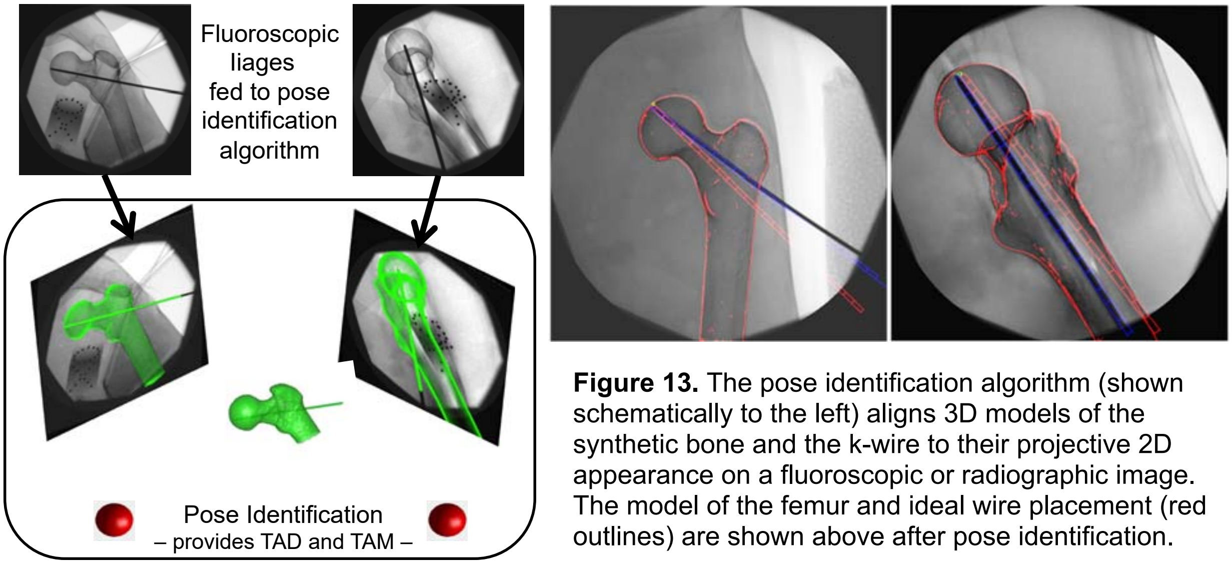

Figure 13. The pose identification algorithm (shown schematically to the left) aligns 3D models of the synthetic bone and the k-wire to their projective 2D appearance on a fluoroscopic or radiographic image. The model of the femur and ideal wire placement (red outlines) are shown above after pose identification.

Summary and Conclusions

The work conducted under this grant fully achieved the first research aim and made significant progress toward the ambitious second aim. The first aim was to demonstrate how deliberate practice improves wire navigation performance. We have demonstrated that the wire navigation simulator leads to faster and more precise placement of surgical wires in several experiments with different residents at different training institutions. Furthermore, we have demonstrated how different training approaches can influence the performance gains enjoyed by residents undergoing this training.

The second research aim was to show that simulator performance correlates with performance in the OR. We have established this aim with respect to a mock operating environment, in which real fluoroscopy was used on plastic bones on a simulated patient. We have also developed several assessment techniques that are more reliable and sensitive than any techniques currently used to measure resident performance in the operating room for the repair of a femoral neck fracture. We have shown that our new measurement techniques are sufficiently sensitive to demonstrate experience levels among residents.

The essential remaining question is to discover whether or not training with a simulator can make a measurable impact on the performance of residents working in the operating room. We look forward to answering this question in the next 5 years, now that we have established the resident pool, the simulator, and the measurement techniques necessary to answer this critical question. Should we succeed, this result will provide a more effective and safe way to train surgeons, reduce the time it takes to perform this common and critical surgery, and ultimately improve the care of patients in the United States and throughout the world.

List of Publications and Products

Peer-Reviewed Papers

1. Kho JY, Johns BD, Thomas GW, et al. A hybrid reality radiation-free simulator for teaching wire navigation skills. J Orthop Trauma. 2015;29(10):e385-390. PMID: 26165262; PMC5125723.
2. Thomas GW, Rojas-Murillo S, Hanley JM, et al. Skill assessment in the interpretation of 3D fracture patterns from radiographs. Iowa Orthop J. 2016;36:1–6. PMID: 27528827; PMC4910797.
3. Taylor L, Thomas GW, Karam MD, et al. Assessing wire navigation performance in the operating room. J Surg Educ 2016;73(5):780–787. PMID: 27184177; PMC5131706.
4. Long S, Thomas GW, Anderson DD. Designing an affordable wire navigation surgical simulator. J Med Device. 2016;10(3): 030921. PMID: 27917254; PMC5129743.
5. Long SA, Thomas GW, Anderson DD. Designing an extensible wire navigation simulation platform. J Med Device (Accepted).
6. Taylor LK, Thomas GW, Karam MD, et al. Developing an objective assessment of surgical performance from operating room video and surgical imagery. IIE Trans Healthc Syst Eng. (in Revision).

M.S. Theses

2016 *Developing and Implementing a Computer Vision Based Surgical Simulator for Hip Wire Navigation*. M.S. Thesis, S.A. Long, Biomedical Engineering, The University of Iowa.

2016 *Objective Measures of Operating Room Wire Navigation Performance*. M.S. Thesis, L.K. Taylor, Biomedical Engineering, The University of Iowa.

2016 *Measuring Hip Fracture Fixation Guide Wire Placement for Performance Assessment in Simulation and the Operating Room*. M.S. Thesis, C.E.Q. Rink, Biomedical Engineering, The University of Iowa.

Conference Abstracts

1. Koehler DM, Thomas GW, Karam MD, et al. Surgical coaching from head-camera video for fluoroscopically guided articular fracture surgery. 128th Annual Meeting of the American Orthopaedic Association, June 24-27, 2015, Providence, RI.
2. Long SA, Thomas GW, Taylor LK, et al. A vision based, hybrid reality, wire navigation simulator. 39th Annual Meeting of the American Society of Biomechanics, August 5-8, 2015, Columbus, OH.
3. Taylor LK, Rink CE, Long SA, et al. Assessing wire navigation performance in treating hip fractures. 39th Annual Meeting of the American Society of Biomechanics, August 5-8, 2015, Columbus, OH.
4. Thomas GW, Taylor L, Long SA, et al. Measuring surgical skill in the operating room for orthopaedic simulator validation. 16th International Meeting on Simulation in Healthcare, January 16-20, 2016, San Diego, CA.
5. Long SA, Taylor L, Rink C, et al. Developing a hybrid reality simulator to train orthopaedic residents in wire navigation. 16th International Meeting on Simulation in Healthcare, January 16-20, 2016, San Diego, CA.
6. Long SA, Thomas GW, Anderson DD. Designing an affordable wire navigation surgical simulator. 15th Annual Design of Medical Devices Conference, April 12-14, 2016, Minneapolis, MN.

7. Long SA, Thomas GW, Anderson DD. Designing an extensible wire navigation simulation platform. 16th Annual Design of Medical Devices Conference, April 11–13, 2017, Minneapolis, Minnesota.
8. Karam MD, Long SA, Marsh JL, et al. An augmented reality simulator improves guide wire navigation skills for first-year residents. 2017 American Orthopaedic Association Annual Leadership Meeting, June 20–24, 2017, Charlotte, North Carolina.
9. Karam MD, Taylor L, Anderson DD, et al. Measures of hip fracture wire navigation performance in the operating room reflect surgical experience. 2017 American Orthopaedic Association Annual Leadership Meeting, June 20–24, 2017, Charlotte, North Carolina.
10. Long S, Thomas G, Chrisman M, et al. Simulator training leads to improved wire navigation in first year orthopaedic residents. 41st Annual Meeting of the American Society of Biomechanics, August 8–11, 2017, Boulder, Colorado.
11. Long SA, Thomas GW, Chrisman M, et al. Simulator training leads to improved wire navigation in first year residents. 64th Annual Meeting of the Orthopaedic Research Society, March 10–13, 2018, New Orleans, Louisiana.

Invited Workshops/Symposia/Lectures

2015	Design of Medical Devices Conference (Anderson DD) *Improving Patient Safety in Orthopaedic Trauma Surgical Training* Minneapolis, MN. April 13-16, 2015.

Concurrently Funded Grants based on this work

2014 – 2016	**The American Board of Orthopaedic Surgery** – Innovations in Resident Surgical Education. Transferability of Wire Navigation Skills Gained Using a Radiation-Free Simulator. $25,000 Total Costs.

Follow-On Funded Grants based on this work

2017-2022	**Agency for Healthcare Research and Quality 1 R18 HS025353-01** Simulation to Support Competency-Based Training in Orthopedic Trauma $1,876,255 Total Costs.

Pending Grants based on this work

2018 – 2019	**Pediatric Orthopaedic Society of North America** The Design and Validation of a Wire Navigation Simulator for Pediatric Supracondylar Humerus Fractures. $49,701 Total Costs.