Published in final edited form as:

J Surg Educ. 2016 ; 73(5): 780–787. doi:10.1016/j.jsurg.2016.03.018.

Assessing Wire Navigation Performance in the Operating Room

Leah Taylor, BS^{a,b}, Geb W. Thomas, PhD^{a,c}, Matthew D. Karam, MD^a, Clarence Kreiter, PhD^d, and Donald D. Anderson, PhD^{a,b,c}

^aDepartment of Orthopedics & Rehabilitation, The University of Iowa, Iowa City, IA, USA

^bDepartment of Biomedical Engineering, The University of Iowa, Iowa City, IA, USA

^cDepartment of Mechanical & Industrial Engineering, The University of Iowa, Iowa City, IA, USA

^dDepartment of Family Medicine, The University of Iowa, Iowa City, IA, USA

Abstract

Objective—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. The study objective is to develop meaningful, objective measures of wire navigation performance in the OR.

Design—Resident surgeons wore a head-mounted video camera while performing surgical open reduction and internal fixation using a dynamic hip screw. Data collected from video included: duration of wire navigation, number of fluoroscopic images, and the degree of intervention by the surgeon's supervisor. To determine reliability of these measurements, four independent raters performed them for two cases. Raters independently measured the tip-apex distance (TAD), which reflects the accuracy of the surgical placement of the wire, on all 7 cases.

Setting—University of Iowa Hospitals and Clinics in Iowa City, IA – a public tertiary academic center.

Participants—Seven surgeries were performed by seven different orthopaedic residents. All ten raters were biomedical engineering graduate students.

Results—The means and standard deviations for AP, lateral, and combined TAD measurements of the 10 raters were 2.7, 1.9, and 3.7 mm, respectively, and inter-rater reliability produced a Cronbach's Alpha of 0.97. The inter-rater reliability analysis for all 9 video-based measures produced a Cronbach's Alpha of 0.99.

Conclusions—Several video-based metrics were consistent across the four video reviewers and are likely to be useful for performance assessment. The TAD measurement was less reliable than previous reports have suggested but remains a valuable metric of performance. Non-experts can reliably measure these values and they offer an objective assessment of OR performance.

Keywords

orthopaedic residency; skill assessment; objective; competency; operating room

Introduction

An orthopedic surgeon must pass a written examination, an oral examination, and complete a 5-year residency program at an accredited institution to become board-certified in the United States. Case logs, required by the Accreditation Council for Graduate Medical Education, track the quantity of and level of participation in specific procedures. These logs are used to establish competence, but fall short of measuring technical skill.¹ Only 31% of orthopaedic resident leaders feel that the case log volume should play a role in credentialing.² It is unclear what might replace case logs, but the president of the American Orthopaedic Association recently argued for competency-based training.³ A program focusing on competency-based end points offers advantages over a time-based system, including increased efficiency for learners and opportunities for focused instruction on problem areas.⁴ A crucial step toward a competency-based program is the identification of objective, quantifiable measures for assessing surgical performance.

Presently, there is no widely accepted tool in clinical practice for efficiently and reliably measuring technical surgical skill.⁵ The Objective Structured Assessment of Technical Skill (OSATS) is a reliable and valid tool that has been increasingly used in orthopaedic skills training. It uses a global rating approach to structure expert evaluation of technical skills, with experts working from a list of specific operative competencies that are each rated on a 5-point Likert scale anchored by behavioral descriptors. The OSATS is widely accepted as the “gold standard” for formative surgical skill assessment, at least partly because of the significant correlation between OSATS scores and postgraduate residency year.⁶ OSATS is not recommended for summative evaluation.

OSATS scoring does have several important limitations including: potential personal bias,⁷ the inconvenience and expense of using staff surgeon experts for grading, and the lack of objective demarcations between acceptable and unacceptable performance.⁸ Furthermore, a recent study showed that OSATS scoring methods do not effectively assess the quality of the surgical result, suggesting that efforts must be made to incorporate assessment metrics that reflect the quality of the surgical result.⁹ A more systematic and efficient approach might include a controlled simulation with truly objective scoring metrics. The Fundamentals of Laparoscopic Surgery trainer, a hands-on skill examination required for certification by the American Board of Surgery, is one such example of this approach. The examination measures technical competence of laparoscopic surgical skills by grading the efficiency and precision of five tasks executed on a box trainer.¹⁰

Other simulators provide quantifiable performance metrics that may eventually be useful in measuring operating room (OR) performance. For example, a promising set of visual parameters: prevalence of instrument loss, triangulation time, and prevalence of lookdowns, appear to correlate with other validated skill assessments, including a global rating scale and motion analysis.¹¹ Motion tracking may also one day prove to be useful in the OR. The

Imperial College surgical assessment device (ICSAD) uses electromagnetic tracking to collect motion data and measure metrics such as number and speed of hand movements, distance traveled, and overall time.¹²

Collecting videos of a surgery presents a more immediate approach to measuring OR performance. To relate surgical technical skill to outcome, Birkmeyer et al. used blinded video recordings of bariatric surgeries. A video recording of an operation was graded in terms of technical skill by blinded surgeons, and this score was correlated to postoperative complication rate. Lower skill was associated with an increase in operation time, as well as with rate of reoperation and complication.¹³ In a study by Beard et al., two different surgeons assessed direct observation in the OR completed by a third surgeon and video recordings of the same operations.¹⁴ Inter-rater reliability between the three assessing surgeons was 0.96, suggesting that with the elimination of bias, it is possible to use video-recorded assessments for summative assessments.

Although there are not yet any studies of measuring *orthopedic* surgical skill with video recordings, hip fracture surgeries are a good place to start because they are common and they demand technical proficiency in wire navigation, a generalizable skill. The Accreditation Council for Graduate Medical Education requires that a minimum of 30 cases be logged during an orthopedic residency, suggesting that hip fracture surgery is already considered to be a foundational part of an orthopaedic surgeon's skill.¹⁵ Wire navigation is a challenging task involving the insertion of a stainless steel pin (K-wire) along a specific trajectory to a position in a bone; it can also serve as a guide for surgical implant placement. This skill, in one form or another, is a fundamental aspect of 7 of the 16 orthopedic milestone cases, with a procedure's success depending on the path and final position of the wire. In a simulated wire navigation task, experienced surgeons more accurately place the guide wire in less time than novice residents.¹⁶

Wire navigation used in the treatment of inter-trochanteric fractures of the hip with a dynamic hip screw (DHS) presents a particularly good initial target for OR performance assessment, because the success of the surgery depends upon a straightforward measure – the tip-apex distance. The tip-apex distance (TAD) is the sum of the distances from the tip of the hip screw to the apex of the femoral head on the antero-posterior (AP) and lateral radiographs. The sum does not represent the physical three-dimensional distance from the wire to femoral head apex, but rather the addition of two two-dimensional measurements on the orthogonal fluoroscopic images (Figure 1).¹⁷ A TAD value of 25 mm or less significantly decreases the risk of screw cut out; therefore, surgeons placing a DHS seek to achieve a center-center position with a TAD of less than 25 mm.¹⁸ Although TAD provides an unambiguous measure of surgical placement, OR performance must also account for how a surgeon balances precision against surgery duration.

Operation time correlates with both cost and patient outcomes. A surgeon's experience and competence tends to reduce operating time.¹⁹ An operation time of greater than 240 minutes carries a significantly higher risk of surgical site infection in hip fractures.²⁰ Surgical performance in orthopedics also depends on the effective use of fluoroscopic imaging. Intraoperative fluoroscopy is crucial to visualization during many orthopedic procedures but

is also a source of radiation exposure to the patient and OR staff. Adverse effects of radiation exposure due to scatter during C-arm fluoroscopic imaging include an increased risk of cancer, sterility, and eye damage.^{21–23} Kahn et al. explored radiation used by residents in DHS procedures and showed the primary differentiator in radiation exposure is surgical experience.²⁴ Multiple studies have also shown that an increase in surgical experience results in a decrease of radiation exposure.^{21, 25–27}

Another measure of surgical performance that considers how meticulous and efficient the surgeon is at exposing the patient to the least possible insult or injury is the number of holes created in the bone cortex. When a wire does not have an appropriate trajectory, it must be removed then reinserted, which increases surgical time and radiation exposure. If the wire is advanced too far it may breach the femoral head and disrupt the joint space.

The principal objective of this study was to quantitatively assess wire navigation skill in the OR. The specific hypothesis is that the tip-apex distance, the surgical duration, the number of fluoroscopic shots, and the number of bone cortex breaches may be consistently measured using augmented video recording of the procedure from a head-mounted camera.

Material and Methods

Surgeries involving open reduction and internal fixation of intertrochanteric hip fractures performed using a DHS were identified and enrolled in the study. Patient and resident informed consent were obtained following the protocol approved by the Institutional Review Board at The University of Iowa, IRB #201409755. The surgeries were completed at the University of Iowa Hospitals and Clinics by the on-call Orthopaedic Trauma team. Patient mean age was 83.7 years (standard deviation 5.6 years), BMI 28.2 (standard deviation 4.1), and 71.4% (5 of 7) were female. One of the operating residents was a PGY-2 (second year of postgraduate training), five were PGY-3, and one was a PGY-5. All raters in the study were biomedical engineering graduate students working in the Orthopedic Biomechanics Laboratory at the University of Iowa.

Resident surgeons wore a GoPro Hero3 (GoPro, San Mateo, CA) head-mounted video camera during the surgery (Figure 2). Recording started before the operating resident entered the sterile field. Intraoperative fluoroscopic images were saved for later evaluation.

The following data were collected by a researcher who viewed the video and saved fluoroscopic images from that portion of the surgery involving the wire navigation procedure, beginning with the first fluoroscope image with the guide wire and ending with the final fluoroscope image with the wire ready for the implant: duration of wire navigation procedure, number of fluoroscopic images collected, number of pullbacks, number of cortex head breaches, number of times the C-arm moved between AP and lateral imaging positions, and the degree of intervention of the surgeon's supervisor. Table 1 provides a summary of the data source for each of these metrics.

The degree of intervention of the surgeon's supervisor was based on tallies of the number of occurrences of each of three different behaviors: 1) supervisor instructions to someone other than the operating surgeon (e.g., the supervising surgeon requesting a fluoroscopic image

from the radiology technician), 2) supervisor instructions to the operating surgeon (e.g., instructions to drill the wire deeper), and 3) the supervisor handling an instrument as more than a helper to the surgeon (e.g., physically controlling the drill). Four raters independently tallied the number of supervisor interventions of each type for two representative cases, based on both the video recording and intraoperative fluoroscopic images. For each case a supervision intervention score was calculated by multiplying the tally of each intervention type with the weights shown in Table 2 and summing the result.

The TAD was measured from DICOM files of the final AP and lateral images of the wire navigation using OsiriX[®] software (Foundation OsiriX, Geneva, Switzerland). Ten raters independently measured the TAD on 7 cases using the protocol described in Johnson et al. (2008). An ellipse was fit to the subchondral bone of the femoral head and a line drawn between the two points where the ellipse met the femoral neck. This line was then bisected with a second perpendicular line, and where that line intersected, the superior portion of the ellipse was used to define the apex.²⁸ Two of the graders repeated the measurements at least 3 weeks after the initial measurements.

Statistical Analysis

The repeatability of the TAD measurement was assessed with the standard deviation and inter-rater reliability of the 10 raters' TAD measurements for each case (210 data points). Individual (AP and lateral) elements of the TAD were evaluated, as well as the total TAD (the sum of the AP and lateral elements). The intra-rater reliability of the TAD measurement was assessed with the two raters' first and second measurements (84 data points). The repeatability of the video metrics was assessed with the standard deviation and inter-rater reliability of the 4 raters' metrics. Intra-rater and inter-rater reliability analysis was assessed with both Cronbach's Alpha and the Intraclass Correlation Coefficient (two-way random model with absolute agreement, 95% confidence). Reliability analysis was done using SPSS Software (version 23, IBM, Armonk, NY). Descriptive statistics were obtained using Minitab (version 17, Minitab, Inc., State College, PA).

Results

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 outlying value (over 3.5 times the average). With this outlying value removed, the average AP standard deviation for case 1 was reduced from 6.0 to 2.5 mm and for all AP cases from 2.6 to 2.1 mm. The sum standard deviation for case 1 is reduced from 6.9 to 3.1 mm and the average for all cases from 3.7 to 3.1 mm.

The inter-rater reliability analysis for all measures together (AP, lateral, and sum) and across the 10 raters produced a Cronbach's Alpha 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 two 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's Alpha and Intraclass Correlation Coefficient (single measures and average measures) were 0.94, 0.89 and 0.94. Rater 2's Cronbach's Alpha and Intraclass Correlation Coefficient (single measures and average measures) were 0.88, 0.79, and 0.88.

Video Metric results

A summary of the metrics obtained from the 4 raters on 2 cases of the surgical video and intraoperative fluoroscopic images can be found in Table 4.

Time (from fluoroscopic timestamps), number of shots, switches between AP and lateral views, and cortex head breaches were consistent for all raters on both cases. Time was also recorded by raters using the video timestamps (results not shown); this approach from estimating duration produced discrepancies between 2 to 21 seconds for case 1 and between 1 to 25 seconds for case 2. For both cases, the number of pullbacks, instructions to others in room, instructions, and supervising surgeon impact score varied between raters. For case 1, the standard deviation for instructions to others, instructions to the operating surgeon, handling of instruments and impact score was 0.6, 1.0, 0.0, and 5.5, respectively. For case 2, the standard deviation for instructions to others, instructions to the operating surgeon, handling of instruments and impact score was 2.4, 2.2, 1.0, and 3.5, respectively. For case 1, 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 9 metrics across both cases produced a Cronbach's Alpha of 0.996 and Intraclass Correlation Coefficient of 0.90 for single measures and 0.99 for average measures. The inter-rater reliability analysis for pullbacks alone for both cases produced a Cronbach's Alpha of 0.33 and Intraclass Correlation Coefficient of 0.95 for single measures and 0.30 for average measures. The inter-rater reliability analysis for the supervising surgeon impact score alone for both cases produced a Cronbach's Alpha of 0.99.

Discussion

The high values of inter-rater reliability (0.969 Cronbach's Alpha) and intra-rater reliability (0.881 and 0.943 Cronbach's Alpha) show that the technique for measuring the TAD from intraoperative fluoroscopic images is valid and reproducible. The magnitude of the TAD sum standard deviation (3.1 mm) for 10 raters across 7 cases seems acceptable given the utility of the measure and keeping in mind that the error is relatively small compared to 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 4 of the 9 metrics. The number of pullbacks was less consistent than some of the other measures, with an inter-rater reliability Cronbach's Alpha 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 where the guide wire breached the cortex of the femoral head limits the study of that parameter, since 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 re-inserted 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.

The timestamps from the fluoroscopic images proved to be a more convenient and consistent measure of surgery duration compared to measuring duration from the video.

Although tallies for each individual supervising surgeon categories varied between raters, the supervising impact score, which is the weighted sum of the three categories, exhibited a Cronbach's Alpha inter-rater reliability measure of 0.99, suggesting a potentially useful new measure of surgeon supervision.

Baumgaertner et al.'s seminal TAD study measured the inter-rater variability with the average TAD standard deviation. Their reported value of 1.7 mm is much smaller than the standard deviation of 3.1 mm observed in this study. The greater variability in the present study may be attributed to using more raters (ten rather than two), or because the raters were less experienced with the measure than those in the Baumgaertner study. Since each rater viewed the same images, the variability is a consequence of differences in matching an ellipse to the femoral head subchondral bone and/or intersection of this ellipse with the femoral neck (Figure 3). Baumgaertner et al. measured the intra-observer with the standard deviation of the averaged sum measure at two different time points. Again their report of 1.2 mm is smaller than the 2.1 mm observed in this study. The difference was clinically relevant: in one of the seven cases, the case-specific inter-observer standard deviation of 2.8 mm was large enough to yield erroneous conclusions on the mean sum measure of 23.7 mm when compared to the 25 mm threshold.

Using edited video of 25–40 minutes in length, Birkmeyer et al. showed that surgical skill assessed from video recordings of operations correlates to complication rate. Their sensitivity analysis showed that the video analysis had low rater bias and showed that repeated ratings of the same surgeon in different videos were highly correlated, with mean skills rating correlation of $R=0.94$.¹³ The Birkmeyer study also demonstrated that lower technical skill is associated with a 40% longer operating time.¹³ The metrics selected in this study are different than those used in the Birkmeyer study because they focus on orthopaedic surgery rather than bariatric surgery, but they show a similar robustness and repeatability. Future studies will reveal whether they have the same consistency across physicians and whether they correlate with other measures of operating room performance.

OSATS, the gold standard measure of surgical performance, has been extensively studied; OR video assessment using the OSATS global rating scale has an inter-rater reliability ranging from 0.57 to 0.83.⁸ Video review utilizing OSATS interval assessment of resident performance had an inter-rater reliability of $91\pm4\%$.²⁹ The Ottawa Score, which uses a 5-point scale assessing readiness for independent practice, had a standard deviation for technical performance score of 1.01.³⁰ In laparoscopic surgery, a 6-point scale with the amount of support required had an inter-rater reliability score of 0.87.³¹ The inter-rater Cronbach's Alpha for the 4 raters on the 2 cases for all video metrics (time, number of shots, pullbacks, cortex head breaches, switch between AP/lateral shots, supervisor instruction to

others, supervisor instructions to surgeon, supervisor handle of instruments, and overall supervisor impact score) was 0.99, which compares quite favorably to these other techniques, suggesting that the techniques proposed may enjoy even greater inter-rater consistency than OSATS.

Limitations

The TAD measurements previously reported were made for the cannulated lag screw that is placed over the guide wire that we used for our measurements. As the guide wire is used to place the implant, we believe the measure provides an accurate representation of the implant position. Lag screws are available in increments of 5 mm, which limits the surgeons' accuracy in achieving an ideal TAD compared to the guide wire. The TAD measure is sensitive to details of the fluoroscope position and the quality of the radiographic image. If the resident alters the wire position after the final AP and lateral shots, this change in position is not accounted for with the TAD measure. The present study relies on a limited number of cases, all from a single hospital. In some cases, videos were occasionally obscured or did not capture all areas of interest, and audio cues were occasionally unclear, which limited uniform measurement. We made no attempt to account for differences among the teaching styles (i.e., proclivity toward intervention) of the supervising surgeons, but instead attempted to measure the degree to which such interventions occurred in any given surgery.

Conclusions

This study included four raters viewing fluoroscopic images and video collected from a head-mounted camera for seven surgeries. The raters independently rated nine metrics from the data. 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 tip-apex distance was less reliable than previous reports have suggested, but is still a valuable metric of surgeon skill. Since non-experts can reliably measure these metrics, they offer an opportunity for objective, consistent assessment of operating room performance.

Acknowledgments

This work was partially supported by grants from the American Board of Orthopaedic Surgery as well as the Agency for Healthcare Research and Quality (Grant #:1 R18 HS022077-01A1). We would like to thank Andres Alfaro for his help in editing and preparing the manuscript.

References

1. Kim MJ, et al. Refining the evaluation of operating room performance. *J Surg Educ.* 2009; 66(6): 352–6. [PubMed: 20142134]
2. Jeray KJ, Frick SL. A survey of resident perspectives on surgical case minimums and the impact on milestones, graduation, credentialing, and preparation for practice. *AOA Critical Issues.* 2014; 96(23):e195.
3. Marsh JL. AOA 2014–2015 Presidential Address: “Tipping Points” in Surgical Education. *J Bone and Joint Surg Am.* 2015; 97-A(3):5.

4. Marsh JL. Should time spent in residency define the end of training? *J Bone and Joint Surg Am.* 2013; 95(21):1905;. [PubMed: 24196458]
5. Aggarwal R, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Ann Surg.* 2007; 245(6):992–999. [PubMed: 17522527]
6. Niitsu H, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today.* 2013; 43(3): 271–275. [PubMed: 22941345]
7. Cornelis J, et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery.* 2014; 156(5):1078–1088;. [PubMed: 25231747]
8. van Hove PD, et al. Objective assessment of technical surgical skills. *Brit J Surg.* 2010; 97(7):972–987. [PubMed: 20632260]
9. Anderson DD, et al. Objective Structured Assessments of Technical Skills (OSATS) does not assess the quality of the surgical result effectively. *Clin Orthop Relat R.* 2015:1–8.
10. Vassiliou MC, et al. FLS and FES: comprehensive models of training and assessment. *Surg Clin North Am.* 2010; 90(3):535–558. [PubMed: 20497825]
11. Alvand A, et al. Simple visual parameters for objective assessment of arthroscopic skill. *J Bone Joint Surg Am.* 2012; 94(13):e97. [PubMed: 22760398]
12. Moorthy K, et al. Objective assessment of technical skills in surgery. *Brit Med J.* 2003; 327(7422): 1032–1037;. [PubMed: 14593041]
13. Birkmeyer JD, et al. Surgical skill and complication rates after bariatric surgery. *New Engl J Med.* 2013; 369(15):1434–1442;. [PubMed: 24106936]
14. Beard JD, et al. Assessing the technical skills of surgical trainees. *Brit J Surg.* 2005; 92(6):778–782. [PubMed: 15810048]
15. Salazar D, et al. Variability in Accreditation Council for Graduate Medical Education resident case log system practices among orthopaedic surgery residents. *J Bone Joint Surg Am.* 2014; 96(3):e22. [PubMed: 24500594]
16. Thomas GW, et al. The validity and reliability of a hybrid reality simulator for wire navigation in orthopedic surgery. *IEEE T Hum-Mach Sys.* 2015; 45(1):119–125.
17. Baumgaertner MR, et al. The value of the tip-apex distance in predicting failure of fixation of peritrochanteric fractures of the hip. *J Bone Joint Surg Am.* 1995; 77(7):1058–1064;. [PubMed: 7608228]
18. Kumar AJ, et al. Significance of hip rotation on measurement of 'Tip Apex Distance' during fixation of extracapsular proximal femoral fractures. *Injury.* 2007; 38(7):792–796. [PubMed: 17296198]
19. Farnworth LR, et al. A comparison of operative times in arthroscopic ACL reconstruction between orthopaedic faculty and residents: the financial impact of orthopaedic surgical training in the operating room. *Iowa Ortho J.* 2001; 21:31–35.
20. Edwards C, et al. Early infection after hip fracture surgery: risk factors, costs and outcome. *J Bone Joint Surg Br.* 2008; 90-B(6):770–777.
21. Giordano BD, et al. Radiation exposure issues in orthopaedics. *J Bone Joint Surg.* 2011; 93(12):e69. [PubMed: 21498490]
22. Maxon HR, et al. Ionizing irradiation and the induction of clinically significant disease in the human thyroid gland. *Am J Med.* 1977; 63(6):967–978. [PubMed: 343589]
23. Hynes DE, et al. Ionising radiation and the orthopaedic surgeon. *J Bone Joint Surg Br.* 1992; 74(3): 332–334. [PubMed: 1587871]
24. Khan IA, et al. Risk of ionising radiation to trainee orthopaedic surgeons. *Acta Orthop Belg.* 2012; 78(1):106–110. [PubMed: 22523936]
25. Palácio EP, et al. Exposure of the surgical team to ionizing radiation during orthopedic surgical procedures. *Revista Brasileira de Ortopedia.* 2014; 49(3):227–232. [PubMed: 26229805]
26. Bar-On E, et al. Intraoperative C-arm radiation affecting factors and reduction by an intervention program. *J Pediatr Orthop.* 2010; 30(4):320–323. [PubMed: 20502229]

27. Blattert TR, et al. Skill dependence of radiation exposure for the orthopaedic surgeon during interlocking nailing of long-bone shaft fractures: a clinical study. *Arch Orthop Trauma Surg.* 2004; 124(10):659–664. [PubMed: 15365718]
28. Johnson LJ, et al. Measuring tip–apex distance using a picture archiving and communication system (PACS). *Injury.* 2008; 39(7):786–790. [PubMed: 18539280]
29. Seymour NE, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 2002; 236(4):458–63. discussion 463–4. [PubMed: 12368674]
30. Gofton WT, et al. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med.* 2012; 87(10):1401–1407;. [PubMed: 22914526]
31. Miskovic D, et al. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. *Surg Endosc.* 2011; 25(4):1136–42. [PubMed: 20835723]

Figure 1.
Sample construction of measuring the TAD on AP (top) and lateral (bottom) fluoroscopic images.

Figure 2.
Intraoperative picture of a resident operating (left) and camera view during wire navigation (right).

Figure 3. Example variation in AP (left) and lateral (right) TAD measurements due to differences in the geometric construction.

Table 1

Summary of Scoring Metric 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	

Table 2

Categorical Weighting of Supervision Intervention

Impact	Weight/Incidence
Instructions to others in room	1
Instructions/tips	2
Handle of Instrument (more than assisting)	3

Table 3

Mean and Standard Deviation in TAD Measurements of 7 Cases by 10 Raters

Case	<u>Mean (standard deviation) (mm)</u>		
	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)

Table 4

Summary of Video Metrics, Categories with Discrepancies are in Italics

Case	Rater	Time (min)	Shots	Switch AP/Lateral	Breach Cortex	Pull Backs	Instruction to Others	Instruction	Handle Instruments	Score
1	1	8.6	22	1	0	1	1	3	0	7
	2	8.6	22	1	0	2	1	3	0	7
	3	8.6	22	1	0	2	0	3	0	6
	4	8.6	22	1	0	2	0	1	0	2
			Mean (std. deviation)			1.8 (0.5)	0.5 (0.6)	2.5 (1.0)	0.0 (0.0)	5.5 (2.4)
2	1	17.1	27	4	0	2	10	15	0	40
	2	17.1	27	4	0	1	9	17	0	43
	3	17.1	27	4	0	3	5	13	2	37
	4	17.1	27	4	0	5	6	18	1	45
			Mean (std. deviation)			2.8 (1.7)	7.5 (2.4)	15.8 (2.2)	0.8 (1.0)	41.3 (3.5)