



Published in final edited form as:

Surg Endosc. 2017 August ; 31(8): 3313–3319. doi:10.1007/s00464-016-5364-1.

Laparoscopic skill assessment of practicing surgeons prior to enrollment in a surgical trial of a new laparoscopic procedure

Benjamin Zendejas¹, James W. Jakub¹, Alicia M. Terando², Amod Sarnaik³, Charlotte E. Ariyan⁴, Mark B. Faries⁵, Sabino Zani Jr.⁶, Heather B. Neuman⁷, Nabil Wasif⁸, Jeffrey M. Farma⁹, Bruce J. Averbook¹⁰, Karl Y. Bilimoria¹¹, Douglas Tyler¹², Mary Sue Brady^{4,11}, and David R. Farley¹

¹Department of Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

²Department of Surgery, Ohio State University Medical Center, Columbus, OH, USA

³Department of Surgery, H. Lee Moffitt Cancer Center, Tampa, FL, USA

⁴Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁵Department of Surgical Oncology, John Wayne Cancer Institute, Santa Monica, CA, USA

⁶Department of Surgery, Duke University School of Medicine, Durham, NC, USA

⁷Division of General Surgery, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

⁸Department of Surgery, Mayo Clinic, Phoenix, AZ, USA

⁹Department of Surgery, Fox Chase Cancer Center, Philadelphia, PA, USA

¹⁰Department of Surgery, MetroHealth Medical Center, Cleveland, OH, USA

¹¹Department of Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

¹²Department of Surgery, University of Texas Medical Branch, Galveston, TX, USA

Abstract

Background—Outcomes of surgical trials hinge on surgeon selection and their underlying expertise. Assessment of expertise is paramount. We investigated whether surgeons' performance measured by the fundamentals of laparoscopic surgery (FLS) assessment program could predict their performance in a surgical trial.

Methods—As part of a prospective multi-institutional study of minimally invasive inguinal lymphadenectomy (MILND) for melanoma, surgical oncologists with no prior MILND experience underwent pre-trial FLS assessment. Surgeons completed MILND training, began enrolling patients, and submitted videos of each MILND case performed. Videos were scored with the

jakub.james@mayo.edu.

Compliance with ethical standards

Disclosures Benjamin Zendejas, James W. Jakub, Alicia M. Terando, Amod Sarnaik, Charlotte E. Ariyan, Mark B. Faries, Sabino Zani, Jr., Heather B. Neuman, Nabil Wasif, Jeffrey M. Farma, Bruce Averbook, Karl Y. Bilimoria, Douglas Tyler, Mary Sue Brady, David R. Farley have no conflicts of interest or financial ties to disclose.

global operative assessment of laparoscopic skills (GOALS) tool. Associations between baseline FLS scores and participant's trial performance metrics were assessed.

Results—Twelve surgeons enrolled patients; their median total baseline FLS score was 332 (range 275–380, max possible 500, passing >270). Participants enrolled 87 patients in the study (median 6 per surgeon, range 1–24), of which 72 (83%) videos were adequate for scoring. Baseline GOALS score was 17.1 (range 9.6–21.2, max possible score 30). Inter-rater reliability was excellent (ICC = 0.85). FLS scores correlated with improved GOALS scores ($r = 0.57$, $p = 0.05$) and with decreased operative time ($r = -0.6$, $p = 0.02$). No associations were found with the degree of patient recruitment ($r = 0.02$, $p = 0.7$), lymph node count ($r = 0.01$, $p = 0.07$), conversion rate ($r = -0.06$, $p = 0.38$) or major complications ($r = -0.14$, $p = 0.6$).

Conclusions—FLS skill assessment of surgeons prior to their enrollment in a surgical trial is feasible. Although better FLS scores predicted improved operative performance and operative time, other trial outcome measures showed no difference. Our findings have implications for the documentation of laparoscopic expertise of surgeons in practice and may allow more appropriate selection of surgeons to participate in clinical trials.

Keywords

Assessment; Laparoscopy; Clinical trial; Simulation; Melanoma; Groin dissection; Minimally invasive; Videoscopic; Inguinal; Learning curve

Randomized controlled trials that compare the effectiveness of one surgical procedure over another depend greatly on comparable levels of expertise among participating surgeons. This concern was raised in the Veterans Affairs cooperative trial that compared open versus laparoscopic mesh-based inguinal hernia repairs, where an inferiority of the laparoscopic approach was encountered [1]. Critics have argued that participating surgeons' inexperience with the novel laparoscopic procedure led to an unfair comparison [2]. In contrast, the COST trial, which compared open versus laparoscopic colectomy for colon cancer and demonstrated non-inferiority for the laparoscopic approach, required participating surgeons to demonstrate procedural competency prior to enrollment in the trial [3].

The ability to assess, in a controlled environment, with a standardized, reliable and valid test, the operative skill of surgeons participating in clinical trials, without placing patients at risk would be invaluable for the design, implementation and success of surgical trials. Unfortunately, the assessment of operative skill is not straight forward. Multiple variables such as judgment, technical proficiency, patient and disease characteristics and operative team dynamics can interplay to affect the outcome of a surgical procedure. Nonetheless, given the recent advances in simulation-based education [4], the assessment of operative skill through simulation has gained increased acceptance [5]. By far, the most studied approach to the assessment of basic laparoscopic skills has been the fundamentals of laparoscopic skill (FLS) program, which relies on five standardized simulation-based tasks to measure surgical skill outside of the operating room [6, 7]. FLS scores have been shown to be reliable and to predict intraoperative performance, both for surgeons in training and for those in practice [8]. In fact, passage of this test is required for graduating general surgery residents before they can sit for written and oral examinations by the American Board of

Surgery. The aim of this study was to determine whether surgeons' performance, as measured by the FLS assessment program, could predict their performance in clinical practice as part of a surgical trial.

Methods

The Safety and Feasibility of Minimally Invasive Inguinal Lymph Node Dissection trial (SAFE-MILND, NCT01500304) is a prospective, multicenter, phase I/II clinical trial in which established melanoma surgeons were trained in a novel procedure, specifically the minimally invasive technique to perform an inguinal lymphadenectomy (MILND) [9, 10]. The primary aim of the trial was to evaluate the safety and feasibility of the MILND in the treatment of melanoma and have been previously published [9, 10]. For this manuscript, we focus on the pre-trial FLS assessment and the relationship of such with certain trial metrics as described below.

Surgeon recruitment, baseline skill assessment and MILND training

A select group of high-volume melanoma surgeons practicing in the USA who were all experienced in the conventional open inguinal lymphadenectomy (performing at least 6 per year), but had no previous MILND experience, were identified via their publication record and reputation. Surgeons were invited to participate in the SAFE-MILND trial and were required to attend an intense, hands-on, oneday training session to learn the innovative MILND procedure. Training sessions were standardized and occurred on two separate dates at Mayo Clinic, Rochester, Minnesota. Complete details of the training intervention have been previously published [9, 10]. Prior to any educational intervention, participants underwent baseline assessment of their basic laparoscopic skills with the FLS assessment program [6]. In brief, the FLS consists of five simulation-based assessment stations that include peg-transfer, circle pattern cutting, ligating loop application, intra-and extracorporeal knot tying. After a brief orientation of each task, participants were allowed a single acclimation attempt followed by one assessment attempt. Standard FLS scoring for certification purposes follows a proprietary system which is not publicly available. We scored the FLS by incorporating time, accuracy and errors as has been previously described by Vassiliou et al. [11]. In addition, although the FLS assessment program is also composed of a knowledge-based test, we elected not to include this in our assessment as our focus was on operative skill; the validity of the knowledge component is not as well established as that of its skill counterpart [8].

After the baseline assessment, attendees watched an instructional video, depicting the MILND procedure in high-quality graphic detail and adhering to effective adult learning principles [12]. The 20-min video included a 3D animated component of the relevant anatomy with key steps of the procedure outlined and visually depicted, as well as a series of edited operative cases, highlighting the critical aspects of the procedure. The video can be seen at <http://medprofvideos.mayoclinic.org/videos/minimally-invasiveinguinal-lymph-node-dissection-milnd>. All attendees then participated in a hands-on cadaveric laboratory in which each participant performed MILND in a controlled environment under instructional supervision, functioning as first assistant for one case and surgeon for one case. Each

participant was given a DVD copy of the educational video to allow them to review the operation in detail as needed after the course.

Video review and assessment of operative performance

After obtaining institutional review board (IRB) approval, each surgeon prospectively enrolled patients and submitted all data to a central site. Participating surgeons were required to video-record all MILND cases performed throughout the study period. Videos were submitted to the central study site, where they were examined for quality and those deemed either incomplete or with poor graphics precluding adequate scoring were discarded. The first (baseline) and last MILND case videos were scored by two independent surgeon raters. Raters had prior MILND experience and laparoscopic expertise. The global operative assessment for laparoscopic surgery (GOALS) tool was used to score the video-recorded MILND operative performances [13]. GOALS is a global rating scale that measures performance in five domains: depth perception, bimanual dexterity, efficiency, tissue handling and autonomy. Each domain is rated on a 5-point Likert scale with descriptive anchors. The degree of case difficulty was assessed also using a 5-point Likert scale as described by Gumbs et al. [14] and incorporated into the GOALS score for a maximum possible score of 30 [15]. GOALS has been shown to demonstrate adequate inter-rater reliability, discriminate between different levels of expertise [16] and changes in GOALS scores after training have been associated with meaningful improvements in patient outcomes [15]. Raters were familiar with the use of the GOALS tool and received a brief pre-rating orientation.

Trial outcome measures

Operative time, lymph node harvest count, need for conversion to open inguinal lymph node dissection and occurrence of any major complication was calculated for each surgeon for their first (baseline) and last MILND case, as well as a respective average for all cases. Complications were graded using the National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE version 4.0). Major complications were defined as those grade 3 (severe or medically significant but not immediately lifethreatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care and activities of daily living) or greater, excluding lymphedema and seroma.

Statistical analyses

Descriptive statistics are provided as counts (percentages), and measures of central tendency, such as mean \pm standard deviation (SD or SEM if standard error of the mean) or median (range or interquartile range [IQR]), are provided based on distribution of data. Associations between baseline FLS scores and GOALS scores, between FLS scores and trial outcomes measures, and between GOALS scores and trial outcome measures were examined with Pearson or Spearman's correlation as appropriate. Strengths of association according to the adjusted correlation coefficients (r) were categorized as negligible ($r = 0.1-0.19$), weak ($r = 0.2-0.29$), moderate ($r = 0.3-0.39$), strong ($r = 0.4-0.69$) or very strong ($r > 0.7$) [17]. Change in GOALS score from first to last MILND case was evaluated with a paired t test. A random sample of 20% of the submitted videos was reviewed and scored in duplicate to calculate inter-rater reliability through intraclass correlation coefficients (ICC). ICC were classified

based on Cohen's classification as large (> 0.8), moderate ($0.5-0.8$) and small/negligible (< 0.5), with greater values indicating better interrater agreement [18]. All hypothesis testing was two-tailed, and a p value of < 0.05 was considered statistically significant. Analyses were performed with JMP software (version 10, SAS Institute Inc., Cary, North Carolina).

Results

Sixteen surgeons from 12 institutions completed the FLS assessment and MILND training program. Thirteen surgeons from 10 institutions opened the IRB-approved trial at their institution and 12 surgeons prospectively enrolled 88 patients to the trial between June 2012 and September 2014. One patient withdrew preoperatively, and 87 operative cases made up the study group. The median MILND procedures performed per participating surgeon were 6 with a range of 1–24.

Median total FLS score was 332 (range 275–380, max possible 500). All participants achieved a passing score (> 270) [19], but only three surgeons (3/12, 25%) demonstrated skills previously shown to be comparable to expert laparoscopic surgeons (> 350 on 500 point scale or 70 on 100 point scale) [16].

We received 79 (91%) video submissions. Seven videos were discarded due to poor quality, yielding 72 (83%) videos adequate for operative performance scoring. Median baseline MILND operative performance (GOALS) score was 17.1 (range 9.6–21.2, max possible score 30). Inter-rater reliability was excellent (ICC = 0.85).

FLS scores correlated with baseline GOALS scores ($r = 0.57$, $p = 0.05$) and with baseline operative time ($r = -0.6$, $p = 0.02$), meaning that greater FLS scores were associated with improved operative performance and shorter operative time (Figs. 1, 2) during the first case. With regard to FLS scores, no associations were found with the degree of patient recruitment ($r = 0.02$, $p = 0.7$), average lymph node count ($r = 0.01$, $p = 0.07$), trial conversion rate ($r = -0.06$, $p = 0.38$) or average major complication rate ($r = -0.14$, $p = 0.6$).

There was no significant improvement in GOALS scores from the first to last MILND video submission (median 4 [IQR 2–8.5, range 0–22] procedures in between), with a mean \pm SEM change in GOALS score of 2.7 ± 2 , $p = 0.2$. Baseline FLS scores did not correlate ($r = 0.36$, $p = 0.34$) with change in GOALS scores from first to last MILND video submission.

Baseline GOALS scores correlated with baseline operative time ($r = -0.6$, $p = 0.04$) and average major complication rate ($r = -0.58$, $p = 0.05$, Fig. 3), meaning that improved operative performance on the first case was associated with fewer major complications over the duration of the trial and shorter operative times. However, GOALS scores (particularly scores from the last case) were not associated with degree of patient recruitment ($r = -0.1$, $p = 0.72$), average lymph node count ($r = 0.01$, $p = 0.9$) or with trial conversion rate ($r = -0.24$, $p = 0.5$).

Discussion

Our study demonstrates that FLS skill assessment of surgeons prior to their enrollment in a surgical trial is feasible. We also demonstrated better FLS scores predicted improved operative performance and decreased operative time; however, other outcome measures showed no difference by FLS performance. Our findings have implications for the recruitment and documentation of laparoscopic expertise of surgeons in clinical trials.

This study adds to the validation argument of FLS as a tool for assessing basic laparoscopic skills of practicing surgeons in a simulated environment [8]. The overall FLS score correlated with outcomes that are clinically relevant such as operative performance and operative time. Other outcome measures such as surgical site infection and lymph node count are likely dependent on multiple factors, and a much larger study would be required to assess whether a participating surgeon's degree of basic laparoscopic skill impact these measures. The finding that GOALS scores were associated with certain trial outcomes also supports the use of such scoring systems for the assessment of laparoscopic skills of practicing surgeons in the clinical setting. Our study results are in line with the results of a recent systematic review of the simulation literature that has shown that simulation-based assessments often correlate with patient outcomes [20]. Nonetheless, the FLS assessment has its limitations. In this study, FLS did not correlate with several important clinical outcomes. This is not unexpected, as FLS scores are more of a proxy for basic laparoscopic skills and would reward speed and efficiency, hence our findings of shorter operative times in those with greater FLS scores. Future research should explore how other assessment instruments, perhaps more procedure specific tools, correlate with trial outcomes that are clinically relevant.

Historically, surgeon recruitment for surgical clinical trials has been akin to convenience sampling, where surgeons are invited to participate based on word of mouth, reputation or their scope of practice, with occasional review of case logs as a form of documenting expertise. Such an approach has compromised the study quality of some surgical clinical trials [1]. On the other hand, the COST trial required participating trial surgeons to submit 20 operative reports and one video-recording of their laparoscopic colectomies to demonstrate competency and adherence to oncologic standards [3]. However, no quantifiable assessment of their operative skill was performed. Being able to objectively quantify and assess the relationship between different degrees of operative skill and trial outcomes would be valuable to the design, implementation and monitoring of quality control measures in any surgical trial. This could be performed as a screening measure prior to trial participation, and if this video review could be performed in real time during the trial, then the potential for early identification of a poor performing surgeon with opportunity for targeted remediation would exist. This would be even more relevant to low volume, complex laparoscopic procedures, such as the MILND described in this report. The downside of this approach is trying to incorporate the success of clinical trials that are dependent on technical skills into broad clinical practice when the only participating surgeons that demonstrated efficacy of the procedure are high-volume individuals with a high degree of experience and technical prowess. Nonetheless, precedent for such efforts exist, the CREST trial which compared carotid endarterectomy versus endovascular stenting established a comprehensive training

and credentialing process to select surgeons and interventionalists [21, 22]. Their rigorous selection process led them to narrow their recruitment from 427 applicants to 224 individuals with the skills required to ensure that the randomized trial results fairly contrasted outcomes between endarterectomy and stenting [21].

Consistent with our previous report of the MILND trial [9], where a significant learning curve was not appreciated, the operative performance scores (GOALS) in this study did not appear to significantly improve over subsequent MILND cases. This could be a reflection of the effective better with specific outcomes. Likewise, the MILND training intervention could have been less effective because of the pre-trial MILND training intervention, a learning curve that has not yet been reached, a wide variation in the number of procedures that each surgeon performed, our analyses were underpowered, or our assessment instruments or endpoints were not sensitive enough to detect a difference if one did exist.

Limitations

Our study has limitations that we acknowledge. Only short-term outcomes were evaluated, and as in any oncologic disease, poor surgical quality demonstrated by greater disease recurrence rates may require larger series and longer time to present itself. We encountered variation in the contribution of cases per surgeon, which could have biased our results to high-volume accruing surgeons; however, this was not possible to standardize and is reflective of clinical practice. Also, it was not possible for us to control the time between baseline FLS skill assessment and the time each surgeon performed their first MILND case; therefore, such time heterogeneity and the potential intervening gap in operative experience of each surgeon could have affected some of our results. In addition, not all participating surgeons were laparoscopically trained during residency or fellowship, and for some melanoma surgeons, laparoscopy makes up a very limited part of their standard practice; as such, diversity in baseline skill could be seen as a bias or strength in our study depending on the lens through which it is viewed. As noted, there are many components of the FLS test and we chose to only compare outcomes with the total FLS score. We did not perform sub-analysis on individual FLS tasks to identify whether one task correlated better with specific outcomes. Likewise, the MILND training intervention could have biased study results by adding additional MILND operative exposure to our surgical cohort between assessment time-points. We did not account for multiple hypothesis testing, and because we did not perform a formal power analysis, there is the possibility that some of our results are a type I statistical error. Nonetheless, we feel that the multi-institutional nature and prospective design of this project add strength to the reproducibility of our study findings.

Implications

Our study provides the first attempt at quantifying the degree of trial surgeons' basic laparoscopic skill set. We have demonstrated the feasibility and potential value of this approach; however, many questions remain unanswered and hence represent opportunity for further research in the field. For example, can the FLS pass/fail score threshold that is currently used as a benchmark for graduating general surgery residents be applied as an inclusion criteria for practicing surgeons being recruited for a surgical trial of a laparoscopic procedure? Would any of the individual FLS tasks be enough to predict certain trial

outcomes and hence simplify the assessment? Can we use the FLS program to train participating surgeons to attain a more uniform performance level and achieve greater in-trial operative performance homogeneity? Can we develop an in-trial video review assessment process linked to a feedback mechanism that would give participating surgeons the opportunity to remediate or perform corrective actions before accruing further patients?

Conclusions

In summary, the assessment of operative skill through simulation has become commonplace in the education of surgical trainees and is beginning to find a role in the assessment of practicing surgeons. Our study results suggest that the assessment of practicing surgeon's skills via FLS, prior to their enrollment in a surgical trial, is feasible and represents a valuable opportunity to potentially improve the quality of surgical clinical trials.

Acknowledgments

We would like to acknowledge Gen Smith who made the success of the trial possible with meticulous study coordination between the numerous sites.

Funding

Funded by the Fraternal Order of Eagles Cancer Research Fund and the Mayo Clinic Department of Surgery Research program. These funding sources had no influence or role in study design; collection, analysis and interpretation of data; writing of the report; or in the decision to submit the article for publication.

Abbreviations

FLS	Fundamentals of laparoscopic surgery
MILND	Minimally invasive inguinal lymphadenectomy
IRB	Institutional review board
GOALS	Global operative assessment of laparoscopic skills
SD	Standard deviation
SEM	Standard error of the mean
ICC	Intraclass correlation coefficients
IRQ	Interquartile range

References

1. Neumayer L, Giobbie-Hurder A, Jonasson O, Fitzgibbons R, Dunlop D, Gibbs J, Reda D, Henderson W (2004) Open mesh versus laparoscopic mesh repair of inguinal hernia. *N Engl J Med* doi:10.1056/NEJM200409303511422
2. Schwaitzberg SD, Jones DB, Grunwaldt L, Rattner DW (2005) Laparoscopic hernia in the light of the Veterans Affairs Cooperative Study 456: more rigorous studies are needed. *Surg Endosc* 19:1288–1289. doi:10.1007/s00464-004-8271-9 [PubMed: 16025196]
3. Fleshman J, Sargent DJ, Green E, Anvari M, Stryker SJ, Beart RW, Hellinger M, Flanagan R, Peters W, Nelson H (2007) Laparoscopic colectomy for cancer is not inferior to open surgery based on 5-

- year data from the COST study group trial. *Ann Surg* 246:655–662. doi:10.1097/SLA.0b013e318155a762 (discussion 662–664) [PubMed: 17893502]
4. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ (2011) Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA J Am Med Assoc* 306:978–988
 5. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R (2013) Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med* 88:872–883. doi:10.1097/ACM.0b013e31828fddcf [PubMed: 23619073]
 6. Fundamentals of Laparoscopic Surgery. <http://www.flsprogram.org/>
 7. Vassiliou MC, Dunkin BJ, Marks JM, Fried GM (2010) FLS and FES: comprehensive models of training and assessment. *Surg Clin North Am* 90:535–558. doi:10.1016/j.suc.2010.02.012 [PubMed: 20497825]
 8. Zendejas B, Ruparel RK, Cook DA (2015) Validity evidence for the fundamentals of laparoscopic surgery (FLS) program as an assessment tool: a systematic review. *Surg Endosc* doi:10.1007/s00464-015-4233-7
 9. Jakub JW, Terando AM, Sarnaik A, Ariyan CE, Faries MB, Zani S, Neuman HB, Wasif N, Farma JM, Averbook BJ, Bilimoria KY, Allred JB, Suman VJ, Grotz TE, Zendejas B, Wayne JD, Tyler DS (2015) Training high-volume melanoma surgeons to perform a novel minimally invasive inguinal lymphadenectomy: report of a prospective multi-institutional trial. *J Am Coll Surg* 222(3):253–260. doi:10.1016/j.jamcollsurg.2015.11.010 [PubMed: 26711792]
 10. Jakub JW, Terando AM, Sarnaik A, Ariyan CE, Faries MB, Zani S, Neuman HB, Wasif N, Farma JM, Averbook BJ, Bilimoria KY, Grotz TE, Allred JBJ, Suman VJ, Brady MS, Tyler D, Wayne JD, Nelson H (2016) Safety and Feasibility of Minimally Invasive Inguinal Lymph Node Dissection in Patients With Melanoma (SAFE-MILND): report of a prospective multi-institutional trial. *Ann Surg* doi:10.1097/SLA.0000000000001670
 11. Vassiliou MC, Ghitulescu GA, Feldman LS, Stanbridge D, Lef-fondré K, Sigman HH, Fried GM (2006) The MISTELS program to measure technical skill in laparoscopic surgery: evidence for reliability. *Surg Endosc* 20:744–747. doi:10.1007/s00464-005-3008-y [PubMed: 16508817]
 12. Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R (2013) Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med Teach* 35:e867–e898. doi:10.3109/0142159X.2012.714886 [PubMed: 22938677]
 13. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190:107–113. doi:10.1016/j.amjsurg.2005.04.004 [PubMed: 15972181]
 14. Gumbs AA, Hogle NJ, Fowler DL (2007) Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg* 204:308–313. doi:10.1016/j.jamcollsurg.2006.11.010 [PubMed: 17254935]
 15. Zendejas B, Cook DA, Bingener J, Huebner M, Dunn WF, Sarr MG, Farley DR (2011) Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. *Ann Surg* 254:502–511 [PubMed: 21865947]
 16. McCluney AL, Vassiliou MC, Kaneva PA, Cao J, Stanbridge DD, Feldman LS, Fried GM (2007) FLS simulator performance predicts intraoperative laparoscopic skill. *Surg Endosc* 21:1991–1995. doi:10.1007/s00464-007-9451-1 [PubMed: 17593434]
 17. Hinkle D, Wiersma W, Jurs S (2003) *Applied statistics for the behavioral sciences*, 5th edn. Houghton Mifflin, Boston
 18. Cohen J (1988) *Statistical power analysis for the behavioral sciences* Routledge Academic, New York
 19. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM (2003) Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. *Surg Endosc* 17:964–967. doi:10.1007/s00464-002-8828-4 [PubMed: 12658417]

20. Brydges R, Hatala R, Zendejas B, Erwin P, Cook D (2015) Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med* 90(2):246–256. doi:10.1097/ACM.0000000000000549 [PubMed: 25374041]
21. Hopkins LN, Roubin GS, Chakhtoura EY, Gray WA, Ferguson RD, Katzen BT, Rosenfield K, Goldstein J, Cutlip DE, Morrish W, Lal BK, Sheffet AJ, Tom M, Hughes S, Voeks J, Kathir K, Meschia JF, Hobson RW, Brott TG (2010) The carotid revascularization endarterectomy versus stenting trial: credentialing of interventionalists and final results of lead-in phase. *J Stroke Cerebrovasc Dis* 19:153–162. doi:10.1016/j.jstrokecerebrovasdis.2010.01.001 [PubMed: 20189092]
22. Hobson RW, Howard VJ, Roubin GS, Ferguson RD, Brott TG, Howard G, Sheffet AJ, Roberts J, Hopkins LN, Moore WS (2004) Credentialing of surgeons as interventionalists for carotid artery stenting: experience from the lead-in phase of CREST. *J Vasc Surg* 40:952–957. doi:10.1016/j.jvs.2004.08.039 [PubMed: 15557910]

FLS Score vs Baseline GOALS Score

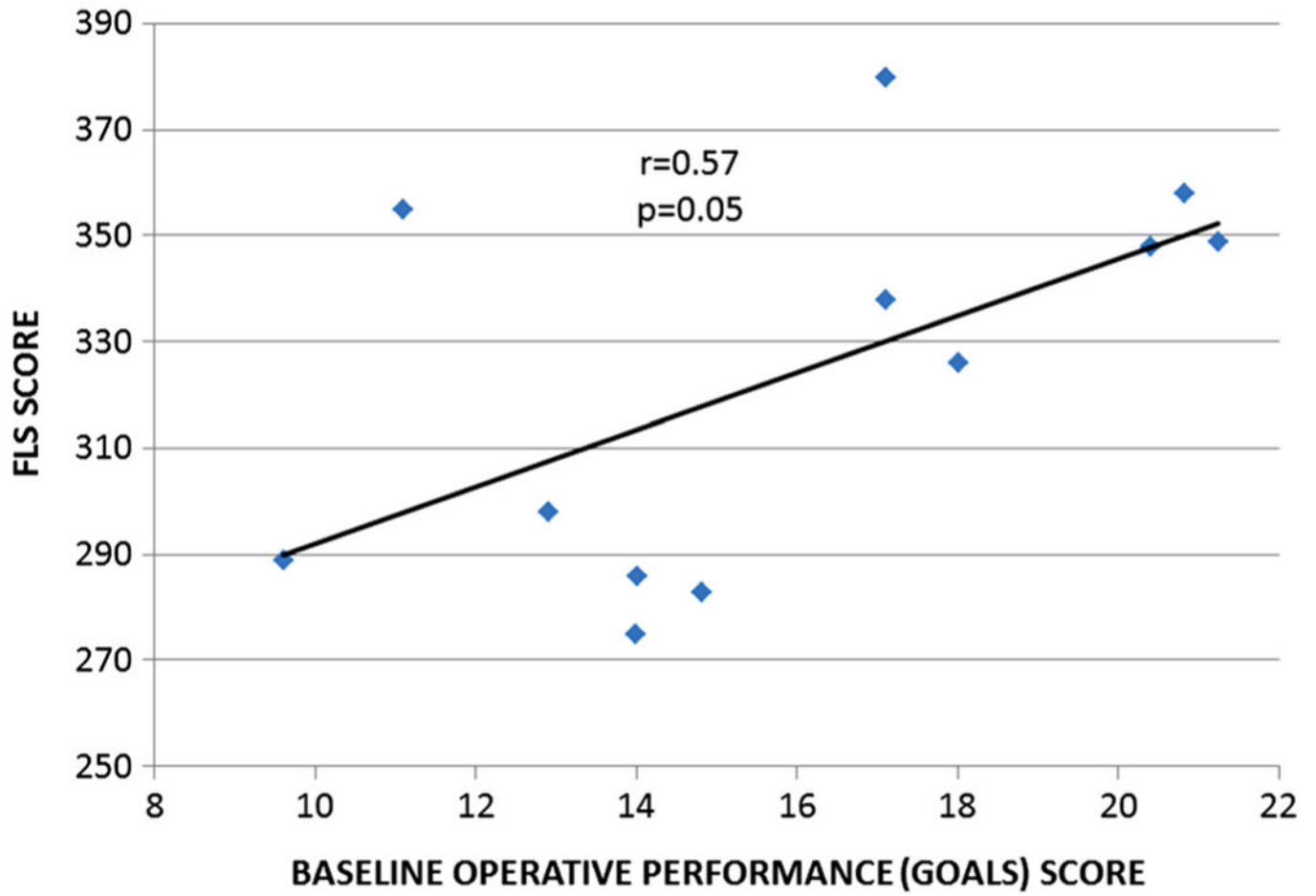


Fig. 1.
Total FLS score versus baseline GOALS score

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FLS vs Baseline Operative Time

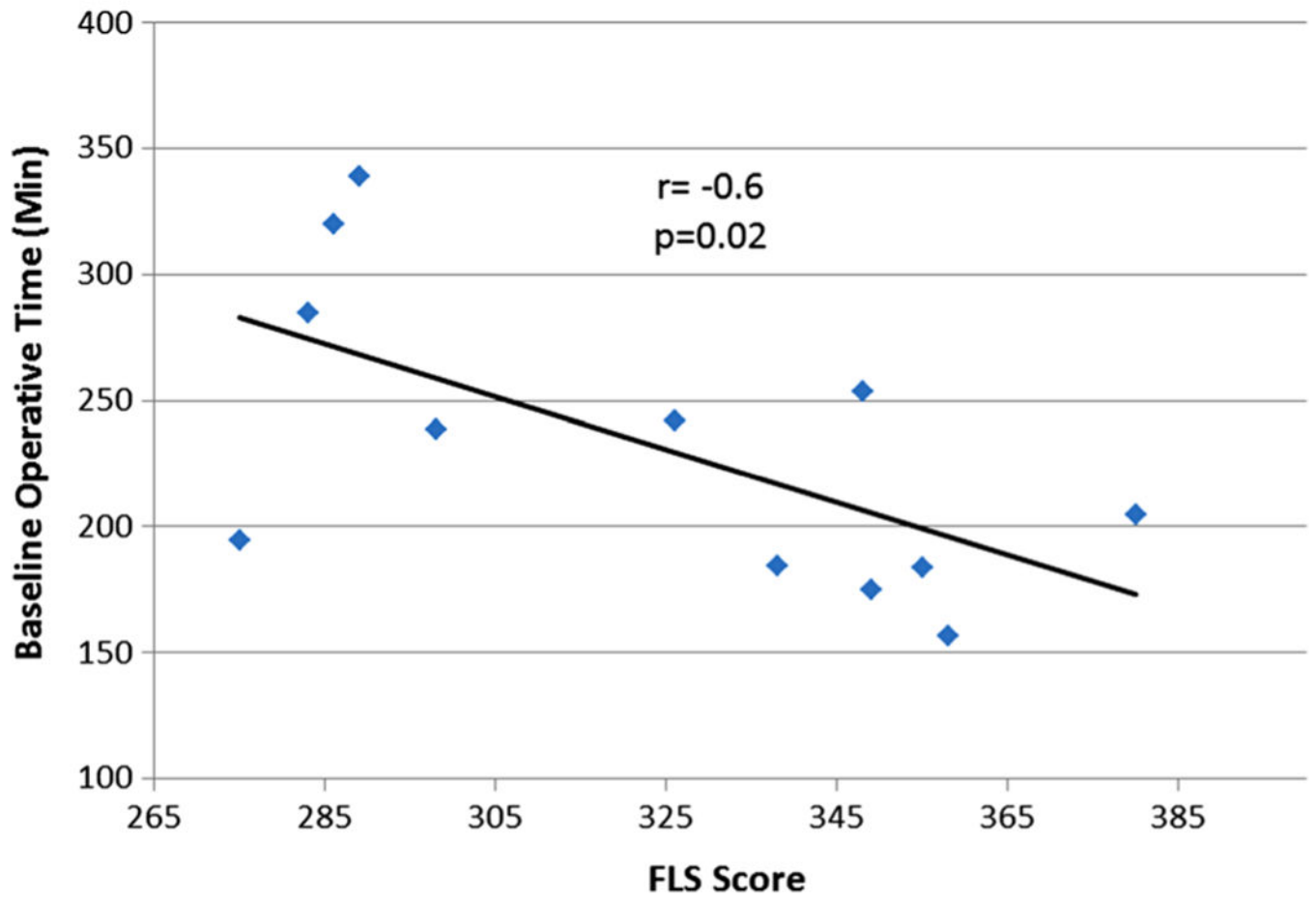


Fig. 2.
Baseline operative time versus total FLS score

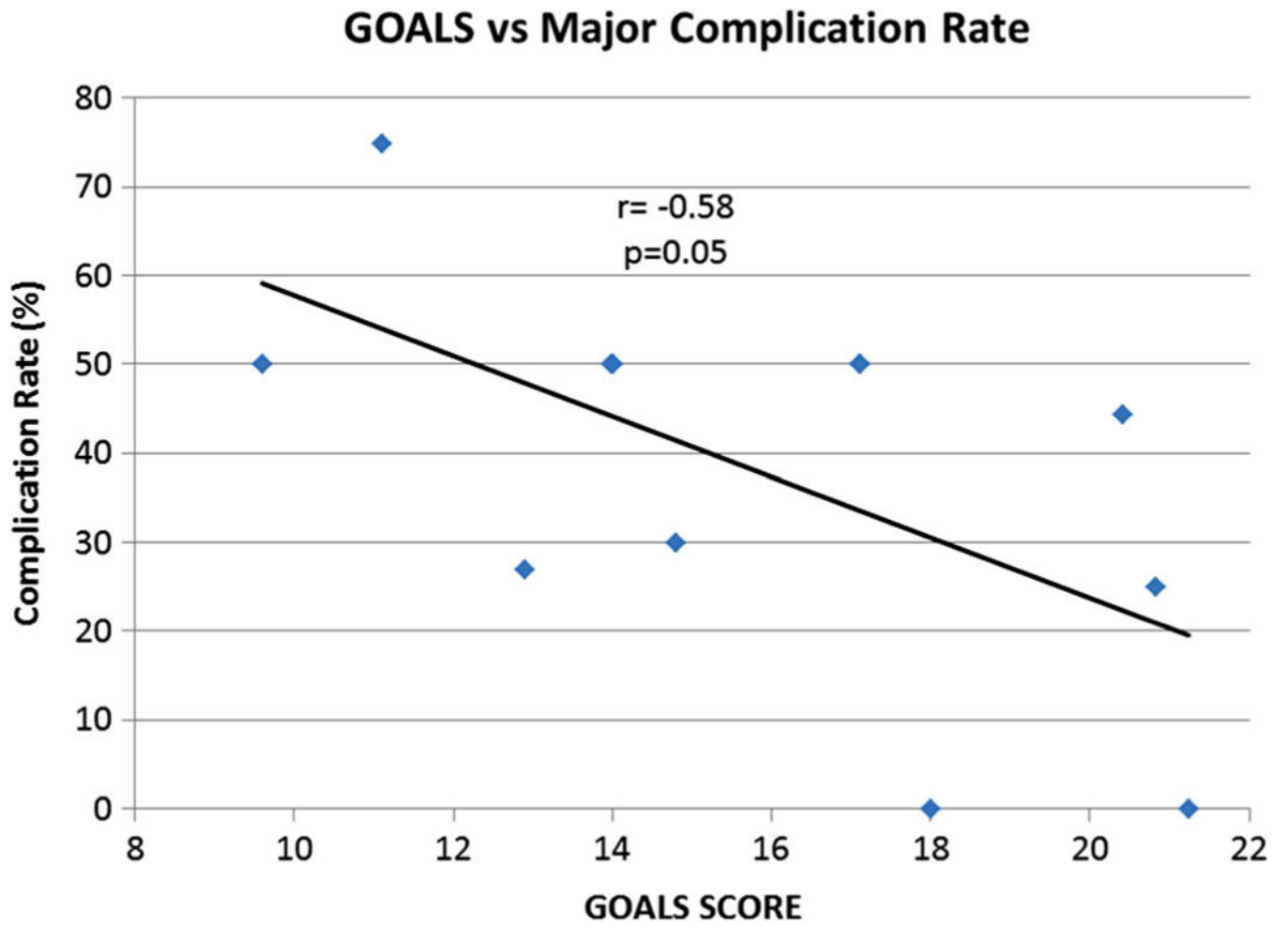


Fig. 3.
Average major complication rate versus baseline GOALS score