



Outcome prediction in bariatric surgery through video-based assessment

Poppy Addison^{1,2} · Daniel Bitner² · Katie Carsky^{1,2} · Saratu Kutana² · Samuel Dechario³ · Anthony Antonacci⁴ · David Mikhail¹ · Samuel Pettit⁵ · Paul J. Chung^{1,6} · Filippo Filicori^{1,2,6}

Received: 19 March 2022 / Accepted: 13 July 2022 / Published online: 4 August 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Introduction The relationship between intraoperative surgical performance scores and patient outcomes has not been demonstrated at a single-case level. The GEARS score is a Likert-based scale that quantifies robotic surgical proficiency in 5 domains. Given that even highly skilled surgeons can have variability in their skill among their cases, we hypothesized that at a patient level, higher surgical skill as determined by the GEARS score will predict individual patient outcomes.

Methods Patients undergoing robotic sleeve gastrectomy between July 2018 and January 2021 at a single-health care system were captured in a prospective database. Bivariate Pearson's correlation was used to compare continuous variables, one-way ANOVA for categorical variables compared with a continuous variable, and chi-square for two categorical variables. Significant variables in the univariable screen were included in a multivariable linear regression model. Two-tailed p -value < 0.05 was considered significant.

Results Of 162 patients included, 9 patients (5.5%) experienced a serious morbidity within 30 days. The average excess weight loss (EWL) was $72 \pm 12\%$ at 6 months and $74 \pm 15\%$ at 12 months. GEARS score was not significantly correlated with EWL at 6 months ($p = 0.349$), 12 months ($p = 0.468$), or serious morbidity ($p = 0.848$) on unadjusted analysis. After adjusting, total GEARS score was not correlated with serious morbidity ($p = 0.914$); however, GEARS score did predict EWL at 6 ($p < 0.001$) and 12 months ($p < 0.001$). All GEARS subcomponent scores, bimanual dexterity, depth perception, efficiency, force sensitivity, and robotic control were predictive of EWL at 6 months ($p < 0.001$) and 12 months ($p < 0.001$) on multivariable analysis.

Conclusion For patients undergoing sleeve gastrectomy, surgical skill as assessed by the GEARS score was correlated with EWL, suggesting that better performance of a sleeve gastrectomy can result in improved postoperative weight loss.

Keywords Video-based assessment · Outcomes · Bariatrics · Robotics

✉ Poppy Addison
paddison@northwell.edu

¹ Department of General Surgery, Lenox Hill Hospital, Northwell Health, 186 E 76th St, 1st Floor, New York, NY 10075, USA

² Intraoperative Performance Analytics Laboratory (IPAL), Department of General Surgery, Lenox Hill Hospital, Northwell Health, New York, NY, USA

³ Institute for Spine and Scoliosis, Lawrenceville, NJ, USA

⁴ Northwell Health, Manhasset, NY, USA

⁵ University of New England College of Osteopathic Medicine, Biddeford, ME, USA

⁶ Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA

While common sense suggests that better skilled surgeons will have better postoperative outcomes, there is surprisingly little literature that tests this hypothesis [1]. Given the constraints of data sharing and patient privacy [2], currently available studies tend to summarize a surgeon's technical skill and correlate that score with their overall outcomes [1,3]. We hypothesized that at a patient level, how well a particular operation is performed will correlate with their postoperative outcomes.

Performance assessment has been traditionally carried out through qualitative judgment and informal observation in the operating room. Quantitative scoring systems such as the Objective Structured Assessment of Technical Skill (OSATS) and Global Evaluative Assessment of Robotic Skills (GEARS) have recently been developed to provide

reproducible assessments of surgical skills [4,5]. These assessment tools use a Likert scale for score components of intraoperative performance, namely depth perception, bimanual dexterity, efficiency, force sensitivity, and tissue handling, with each domain scored out of 5 for a total of 25 possible points. The GEARS score is internally validated [6] and is consistent across expert- and crowd-sourced review, allowing laypeople to quantify surgical skill to avoid the costly and time-consuming process of expert review [7,8].

Bariatric surgery uniquely offers standardized procedures, barring some nuance [9], in a relatively healthy patient population with a unique outcome, excess weight loss (EWL). We sought to determine the relationship between postoperative outcomes and intraoperative technical skills for robotic sleeve gastrectomy as quantified by the GEARS score during crowd-sourced video-based assessment.

Methods

Patients undergoing robotic sleeve gastrectomy between July 2018 and January 2021 at a single health care system were captured in a prospective database for retrospective analysis. Given the inability to assign GEARS scores for laparoscopic or open cases, any patient who was converted from a robotic approach was not included. Patients younger than 18 years old were also excluded. GEARS scores were assigned through crowd-sourced evaluators by a third party; the methodology has been previously described by this group [10,11]. Patient identifying information is captured and encrypted with a one-way hashing algorithm. This information and the operative videos are uploaded onto a secure database for assignment of GEARS scores by crowd-source evaluators, which is managed by Crowd-Sourced Assessment of Technical Skills (C-SATS, Seattle, WA). Online evaluators do not have access to any identifying information. Evaluators are trained on VBA and are frequently evaluated against other layperson evaluators and expert surgeon reviewers to determine the reliability of their scoring. After technical skills are assessed by a minimum of 30 evaluators, the scores and hashed identifying number were returned to the research team via a secure application program interface for de-encryption and correlation with patient variables. All data were stored in a secure, HIPAA-compliant database within the surgical department's quality improvement initiative.

Serious morbidity included wound dehiscence, stroke or transient ischemic attack, cardiac arrest, myocardial infarction, pulmonary embolism, deep venous thrombosis, acute kidney injury, sepsis or septic shock, surgical site infection, pneumonia, unplanned intubation, urinary tract infection, ileus, anastomotic or staple line leak, and postoperative

hernia. Complications were only recorded within 30 days of surgery.

Bivariate Pearson's correlation was used to compare continuous variables, one-way ANOVA for categorical variables compared with a continuous variable, and chi-square for two categorical variables. Significant variables in the univariable screen (age, BMI, CCI and ASA) were included in a multivariable linear regression model. Patients lost to follow-up were censored at their last known visit date. Separate models were created for EWL at 6 and 12 months, and each GEARS subcomponent was evaluated in a separate model. No multivariable regression was performed for serious morbidity as there were no significant variables in the univariable screen. Assumptions of linear regression were tested as follows. There is a linear relationship between the outcome variable (excess weight loss) and the independent variables. The independent variables were not highly correlated with each other. All residuals are normally distributed. All analyses were performed with SPSS 26.0 (IBM, Armonk, NY) statistical software. Two-tailed p -value < 0.05 was considered significant. This study was approved by the Institutional Review Board at Northwell Health. Written consent was not required.

Results

A total of 162 patients who underwent robotic sleeve gastrectomy performed by a total of 7 surgeons were captured (Table 1). No patients met exclusion criteria. The majority of patients were young and healthy, with a mean age of 40.8 ± 12.6 years, a mean Charlson comorbidity index (CCI) 0.69 ± 1.2 , and a mean American Society of Anesthesiologists (ASA) score 2.5 ± 0.5 . Most patients were non-Hispanic (73.4%), women (80.2%), split among white (32.6%), Black (25.3%), and other (32.7%) racial identities. From a mean starting BMI of 42.4 ± 5.1 , the mean EWL at 6 months was $72 \pm 11.7\%$ and at 12 months $74.7 \pm 14.5\%$. EWL at 6 months was only available for 88 patients and at 12 months for 55 patients. The mean GEARS score was 20.2 with a standard deviation of 0.72. Mean subcomponent scores were bimanual dexterity 4.1 ± 0.2 ; depth perception 4.0 ± 0.2 ; efficiency 3.8 ± 0.2 ; force sensitivity 4.2 ± 0.2 ; and robotic control 4.2 ± 0.2 . Only 9 patients (5.5%) experienced a serious morbidity, which included 1 patient with a urinary tract infection, 1 pneumonia, 1 acute kidney injury, 2 deep venous thromboses, 2 surgical site infections (1 requiring return to the operating room for washout), and 2 port site hernias.

To further evaluate the potential for confounding, age, sex, race, BMI, and ASA were evaluated on a univariate screen and found to correlate with EWL at 6 months and age,

Table 1 Patient demographics and correlation with excess weight loss

Variable	Mean \pm SD	EWL at 6 months* <i>p</i> -value	EWL at 1 year** <i>p</i> -value
Age (years)	40.8 \pm 12.6	0.008 [†]	0.020 [†]
Sex	80.2% women, 19.8% men	0.034 ^{††}	0.770 ^{††}
Ethnicity	26.6% Hispanic, 73.4% non-Hispanic	0.090 ^{††}	0.538 ^{††}
Race	34.6% White, 32.7% Other, 25.3% Black, 1.9% Multi-racial, 0.6% Asian, 4.9% Not Reported	0.008 ^{††}	0.020 ^{††}
BMI (kg/m ²)	42.4 \pm 5.1	<0.001 [†]	<0.001 [†]
Charlson comorbidity index	0.69 \pm 1.2	0.092 [†]	0.022 [†]
ASA	2.5 \pm 0.5	0.013 [†]	0.320 [†]
Operative time (min)	116 \pm 36	0.130 [†]	0.497 [†]
Estimated blood loss (mL)	18 \pm 0.7	0.572 [†]	0.488 [†]
Length of stay (days)	1.5 \pm 0.7	x	x
BMI 6-month postop (kg/m ²)*	32.1 \pm 4.5	x	x
BMI 1-year postop (kg/m ²)**	30.1 \pm 5.8	x	x
EWL 6 months (%)*	72.0 \pm 11.7	x	x
EWL 1 year (%)**	74.7 \pm 14.5	x	x

n* = 88*n* = 55[†]Bivariate Pearson's correlation^{††}One-way ANOVA**Table 2** Patient demographics and correlation with GEARS score

Variable	Correlation coefficient	<i>p</i> -value
Age (years)	− 0.170*	0.031 [†]
Sex	0.062**	0.435 ^{††}
Ethnicity	− 0.038**	0.637 ^{††}
Race	− 0.117**	0.149 ^{††}
BMI (kg/m ²)	− 0.054*	0.496 [†]
Charlson comorbidity index	− 0.055*	0.486 [†]
ASA	0.108*	0.175 [†]
Operative time (mins)	− 0.142*	0.083 [†]
Estimated blood loss (mL)	− 0.188*	0.017 [†]
Length of stay (days)	− 0.085*	0.283 [†]

*Pearson correlation

**Standardized coefficients

[†]Bivariate Pearson's correlation^{††}One-way ANOVA

race, BMI, and CCI at 12 months (Table 1). The correlation between GEARS score and demographics and outcomes was similarly evaluated (Table 2). The overall GEARS score was correlated with age ($p = 0.031$) and estimated blood loss ($p = 0.017$); however, there was no correlation identified for other patient demographics, including BMI ($p = 0.496$) or outcomes.

The total GEARS score or its subcomponents were not correlated with EWL at 6 or 12 months on unadjusted analysis (Table 3). However, after adjusting for age, sex, race, BMI, CCI, and ASA, total GEARS score and its subcomponents were positively correlated with EWL at 6 and 12 months ($p < 0.001$). There was insufficient evidence to conclude a correlation exists with any patient demographic or GEARS scores and serious morbidity (Table 4).

Discussion

This study evaluated the relationship between intraoperative technical skill and postoperative outcomes for robotic sleeve gastrectomy. We determined that skill as determined by blinded video-based review and quantified with the GEARS score correlates with weight loss. While the overall low number of serious complications would require a much larger study to determine the relationship between skill and serious complications, this work is among few studies that demonstrate that more technically skilled surgeons may have better outcomes. These conclusions have meaningful consequences for surgical credentialing and residency education [2,12].

This is the first study to correlate technical skills of the surgeon with patient outcomes on a patient level. Previous studies asked surgeons to submit a small number of

Table 3 GEARS score and correlation with excess weight loss

Variable	Mean \pm SD	EWL at 6 months**			EWL at 1 year***		
		Unadjusted <i>p</i> -value	Regression coefficient	Adjusted <i>p</i> -value*	Unadjusted <i>p</i> -value	Regression coefficient	Adjusted <i>p</i> -value*
GEARS score	20.2 \pm 0.72	0.349 [†]	0.021	<0.001	0.468 [†]	0.104	<0.001
Bimanual dexterity	4.1 \pm 0.2	0.903 [†]	0.071	<0.001	0.348 [†]	0.071	<0.001
Depth perception	4.0 \pm 0.2	0.190 [†]	0.034	<0.001	0.951 [†]	0.072	<0.001
Efficiency	3.8 \pm 0.2	0.491 [†]	0.058	<0.001	0.870 [†]	0.031	<0.001
Force sensitivity	4.2 \pm 0.2	0.177 [†]	0.037	<0.001	0.598 [†]	0.148	<0.001
Robotic control	4.2 \pm 0.2	0.960 [†]	0.061	<0.001	0.195 [†]	0.126	<0.001

*Controlled for age, sex, race, BMI, Charlson comorbidity index, and ASA

***n* = 88

****n* = 55

[†] Bivariate Pearson's correlation

Table 4 Correlation of patient demographics and GEARS score with serious morbidity

Variable	Standardized correlation coefficient	<i>p</i> -value
Age (years)	0.032	0.859 [†]
Sex	−0.009	0.892 ^{††}
Ethnicity	0.117	0.143 ^{††}
Race	−0.019	0.988 ^{††}
BMI	0.186	0.553 [†]
Charlson comorbidity index	−0.115	0.713 [†]
ASA	0.344	0.273 [†]
Operative time (mins)	−0.278	0.377 [†]
Estimated blood loss (mL)	−0.043	0.891 [†]
GEARS score	0.060	0.848 [†]
Bimanual dexterity	0.084	0.788 [†]
Depth perception	0.106	0.735 [†]
Efficiency	−0.125	0.685 [†]
Force sensitivity	−0.095	0.762 [†]
Robotic control	0.261	0.404 [†]

[†] One-way ANOVA

^{††} Chi-square test

representative intraoperative videos, summarize specific surgeon's skills with one number, and then correlate that skill evaluation with their overall outcomes [3,9,13,14]. In comparison, we correlate individual patient outcomes with the skill demonstrated in their specific surgery. We were able to accomplish this with our encrypted program interface that allows us to share hashed patient identifiers with a separate team for skill evaluation while maintaining patient privacy

[2]. We demonstrated that even for experienced, fellowship-trained bariatric surgeons, the skill with which a particular surgery is performed will impact that specific patient.

The seminal paper by Birkmeyer et al. first established the relationship between postoperative outcomes and technical skill as measured by direct assessment with blinded video review [3]. Since then, numerous studies have sought to replicate these results or expand them into other operations, beyond the original gastric bypass [1]. However, there are few studies that use direct objective measurements of skill rather than a proxy, such as operative time or surgeon experience [1]. Importantly, these proxies have not been validated as a measure of technical skill. Operative time, surgeon experience, length of stay, and complication rates have complex interdependent relationships [1,10,15]. Furthermore, this group asserts that operative time and length of stay are outcomes of skill rather than an indirect measurement of skill itself.

In bariatric surgery, Birkmeyer et al. evaluated 20 surgeons performing laparoscopic gastric bypass and found surgeons at the top quartile of skill had lower complication rates and mortality [3], and similarly, Varban et al. evaluated 25 surgeons performing laparoscopic sleeve gastrectomy and found that more skilled surgeons had lower rates of specific surgical complications but not a lower rate of overall 30-day complications [9]. In robotic surgery, postoperative outcomes have not been correlated with objective technical skill outside of urologic procedures [16,17].

While representing early work in the field of video-based assessment in robotic surgery, this work is not without several important limitations. All surgeons included are fellowship-trained bariatric surgeons operating within a bariatric center

of excellence. This high level of expertise allows us to conclude that even skilled surgeons have small variations case by case that impact patient outcomes. However, it limits the number of patients experiencing complications, precluding our ability to create a regression model that is not over-fit to the data. Our conclusions are also limited to this population of highly experienced surgeons at a bariatric center of excellence; however, robotic bariatric surgery is typically performed in this setting. This highly trained cohort helps explain the small standard deviation of GEARS scores. Additionally, assessment by such a large number of evaluators may have a tendency toward the mean, where small differences in the GEARS score correlates with a large difference in operative skill. While there is trainee involvement in these cases, currently our VBA is limited in that it does not account for which console and therefore which surgeon is performing these cases. When evaluating regression coefficients, the GEARS score and its subcomponents were all positively correlated with EWL; however, the effect sizes were relatively small. Further studies with a larger population and more surgeons across a wider variety of skill may result in a larger effect size.

Our study may also be limited by selection bias. At this institution, we routinely send all robotic bariatric surgical videos for objective scoring. Some videos may not be recorded in their entirety and correlated with patient identifiers, either due to surgeon preference, or technical or human errors. We lost 46% of patients at 6-month ($n=88$) and 66% at 12-month ($n=55$) follow-ups. Compared to other studies of sleeve gastrectomy, the rate lost to follow-up is similar [9]; additionally, our EWL is comparable to that generally reported for sleeve gastrectomy [18]. Finally, while the GEARS score is a validated measure for surgical skill, there is no standard measurement of robotic technical skill [6,7]. The GEARS score was designed to describe the fundamental elements of robotic surgery regardless of the specific procedure [5]. There are numerous other scoring systems that describe nuances of robotic skill, such as specific for microsurgery or control of the console [19,20]. By utilizing the GEARS score, this study can be repeated for any robotic procedure and the results compared across specialties.

This study is also limited in answering the following question: what is a more highly skilled surgeon doing differently than a less skilled surgeon that may result in better weight loss for their patients? To answer this question on a larger scale, this group is looking at kinematic data to break down specific movements. For example, does the angle the stapler takes at the angle of His differ consistently for patients with better EWL? Does more gentle tissue handling result in less swelling of the sleeve and better postoperative outcomes? VBA has been combined with kinematic data derived from the da Vinci system to evaluate robotic performance in other specialties, and our group will next look into applying this data to bariatric surgery [21].

Conclusion

In this retrospective review of patients undergoing robotic sleeve gastrectomy, higher technical skill as assessed by crowd-sourced assignment of the GEARS score did not correlate with serious morbidity but did correlate with weight loss; patients whose cases were assigned a higher GEARS score had more weight loss. Objective, video-based assessment of technical skill may predict postoperative weight loss in robotic sleeve gastrectomy at the patient level.

Funding This work was supported by a 2020 SAGES Robotic Surgery Grant and a 2021 Hugoton Foundation grant.

Declarations

Disclosures Drs. Addison and Bitner are consultants for Deep Surgery. Dr. Filicori is a consultant for Cambridge Medical Robotics, Active Surgical, Boston Scientific, and Digital Surgery. Drs. Carsky, Antonacci, Mikhail, and Chung and Ms. Kutana, Mr. Dechario, and Mr. Pettit have no disclosures.

References

1. Fecso AB, Szasz P, Kerezov G, Grantcharov TP (2017) The effect of technical performance on patient outcomes in surgery: a systematic review. *Ann Surg* 265(3):492–501
2. Filicori F, Addison P (2021) Intellectual property and data ownership in the age of video recording in the operating room. *Surg Endosc* 36(6):3772–3774
3. Birkmeyer JD, Finks JF, O'Reilly A et al (2013) Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 369(15):1434–1442
4. Martin J, Regehr G, Reznick R et al (1995) An objective structured assessment of technical skill (OSATS) for surgical residents. *Gastroenterology* 108(4):A1231
5. Goh A, Goldfarb D, Sander J, Miles B, Dunkin B (2012) Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol* 187(1):247–252
6. Aghazadeh MA, Jayaratna IS, Hung AJ et al (2015) External validation of global evaluative assessment of robotic skills (GEARS). *Surg Endosc* 29(11):3261–3266
7. White L, Kowalewski T, Dockter R, Comstock B, Hannaford B, Lendvay T (2015) Crowd-sourced assessment of technical skill: a valid method for discriminating basic robotic surgery skills. *J Endourol* 29(11):1295–1301
8. Goldenberg MG, Nabhani J, Wallis CJD et al (2017) Feasibility of expert and crowd-sourced review of intraoperative video for quality improvement of intracorporeal urinary diversion during robotic radical cystectomy. *Can Urol Assoc J* 11(10):331–336
9. Varban OA, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Dimick JB (2021) Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy: a video-based study. *Ann Surg* 273(4):766–771
10. Addison P, Yoo A, Duarte-Ramos J et al (2020) Correlation between operative time and crowd-sourced skills assessment for robotic bariatric surgery. *Surg Endosc* 35(9):5303–5309

11. Addison P, Bitner D, Chung P et al (2022) Blinded intraoperative skill evaluations avoid gender-based bias. *Surg Endosc*. <https://doi.org/10.1007/s00464-022-09142-9>
12. McQueen S, McKinnon V, VanderBeek L, McCarthy C, Sonnada R (2019) Video-based assessment in surgical education: a scoping review. *J Surg Educ* 76(6):1645–1654
13. Stulberg JJ, Huang R, Kreutzer L et al (2020) Association between surgeon technical skills and patient outcomes. *JAMA Surg* 155(10):960–968
14. Varban OA, Greenberg CC, Schram J et al (2016) Surgical skill in bariatric surgery: does skill in one procedure predict outcomes for another? *Surg* 160(5):1172–1181
15. Moorthy K, Munz Y, Sarker SK, Darzi A (2003) Objective assessment of technical skills in surgery. *BMJ* 327(7422):1032–1037
16. Goldenberg MG, Goldenberg L, Grantcharov TP (2017) Surgeon performance predicts early continence after robot-assisted radical prostatectomy. *J Endourol* 31(9):858–863
17. Stern J, Sharma S, Mendoza P et al (2011) Surgeon perception is not a good predictor of peri-operative outcomes in robot-assisted radical prostatectomy. *J Robot Surg* 5(4):283–288
18. Peterli R, Wölnerhanssen BK, Peters T et al (2018) Effect of laparoscopic sleeve gastrectomy vs laparoscopic Roux-en-Y gastric bypass on weight loss in patients with morbid obesity: the SM-BOSS randomized clinical trial. *JAMA* 319(3):255–265
19. Chen J, Cheng N, Cacciamani G et al (2019) Objective assessment of robotic surgical technical skill: a systematic review. *J Urol* 201(3):461–469
20. Brown KC, Bhattacharyya KD, Kulason S, Zia A, Jarc A (2020) How to bring surgery to the next level: interpretable skills assessment in robotic-assisted surgery. *Visc Med* 36(6):463–470
21. Lyman WB, Passeri MJ, Murphy K et al (2021) An objective approach to evaluate novice robotic surgeons using a combination of kinematics and stepwise cumulative sum (CUSUM) analyses. *Surg Endosc* 35(6):2765–2772

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.