Surgical Endoscopy (2024) 38:5319–5330
https://doi.org/10.1007/s00464-024-11029-w

2024 SAGES ORAL

Validity of task-specific metrics for assessment in perineal proctectomy

Alexis Desir[1] · Poya Pourghaderi[1] · Shruti R. Hegde[1] · Doga Demirel[2] · Javier Salgado Pogacnik[1] · Suvranu De[3] · James W. Fleshman[4] · Ganesh Sankaranarayanan[1]

Received: 16 April 2024 / Accepted: 30 June 2024 / Published online: 18 July 2024

Abstract

Background Perineal proctectomy is a complex procedure that requires advanced skills. Currently, there are no simulators for training in this procedure. As part of our objective of developing a virtual reality simulator, our goal was to develop and validate task-specific metrics for the assessment of performance for this procedure. We conducted a three-phase study to establish task-specific metrics, obtain expert consensus on the appropriateness of the developed metrics, and establish the discriminant validity of the developed metrics.

Methods In phase I, we utilized hierarchical task analysis to formulate the metrics. In phase II, a survey involving expert colorectal surgeons determined the significance of the developed metrics. Phase III was aimed at establishing the discriminant validity for novices (PGY1-3) and experts (PGY4-5 and faculty). They performed a perineal proctectomy on a rectal prolapse model. Video recordings were independently assessed by two raters using global ratings and task-specific metrics for the procedure. Total scores for both metrics were computed and analyzed using the Kruskal–Wallis test. A Mann–Whitney U test with Benjamini–Hochberg correction was used to evaluate between-group differences. Spearman's rank correlation coefficient was computed to assess the correlation between global and task-specific scores.

Results In phase II, a total of 23 colorectal surgeons were recruited and consensus was obtained on all the task-specific metrics. In phase III, participants ($n=22$) included novices ($n=15$) and experts ($n=7$). There was a strong positive correlation between the global and task-specific scores (rs = 0.86; $P<0.001$). Significant between-group differences were detected for both global ($\chi^2=15.38$; $P<0.001$; df = 2) and task-specific ($\chi^2=11.38$; $P=0.003$; df = 2) scores.

Conclusions Using a biotissue rectal prolapse model, this study documented high IRR and significant discriminant validity evidence in support of video-based assessment using task-specific metrics.

Keywords Simulation · Perineal proctectomy · Task-specific metrics

Current board certification requires trainees of various surgical specialties, including colon and rectal surgery, to pass oral and written board exams to assess their knowledge for competent surgical practice. Despite the well-established fact that surgeon technical skill proficiency significantly impacts patient surgical outcomes and complication rates [1], there is no formal examination of the technical skills of surgical trainees before their transition into independent surgical practice. The sole evaluation of technical proficiency depends on subjective evaluations by faculty in individual training programs that lack objective measures [2, 3]. Studies comparing these subjective assessments to more objective metrics, such as task-specific evaluations, have revealed shortcomings in accurately evaluating a resident's technical performance [4, 5].

Due to this concern, the Operative Assessment Committee of the American Society of Colon and Rectal

Accepted as a podium presentation for the 2024 annual meeting of the Society of American Gastrointestinal and Endoscopic Surgeons in Cleveland, Ohio.

✉ Ganesh Sankaranarayanan
Ganesh.Sankaranarayanan@utsouthwestern.edu

1 Department of Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA

2 Florida Polytechnic University, Lakeland, FL, USA

3 Florida A&M University-Florida State University College of Engineering, Tallahassee, FL, USA

4 Baylor University Medical Center, Dallas, TX, USA

Surgeons introduced the Colorectal Objective Structured Assessment of Technical Skill (COSATS) to address the assessment of technical competency in colorectal surgery trainees. The committee incorporated both a global rating scale and a task-specific checklist for specific tasks deemed fundamental to independent colorectal surgical practice. Both scales successfully differentiated colorectal surgery residents from their general surgery counterparts [6]. We have taken this work and applied hierarchical task analysis to deconstruct the surgical procedures into intricate steps and then crafted evaluation metrics, successfully demonstrating their validity for cricothyrotomy [7], the creation of a double-layered hand-sewn anastomosis in the small bowel [8], and a linear stapler-based small bowel anastomosis [9]. Although these extensive subjective tools have yielded great results, the utilization of skill assessment tools entails substantial expenses for examiner and staff time, labor, and materials for individual participant setups. Furthermore, the need for a proficient proctor to oversee and evaluate the performance introduces an additional obstacle to participation.

As a potential alternative, surgical simulators based on virtual reality offer several advantages, including objective scoring, rapid turnaround times, and the absence of the need to replenish materials. The use of virtual reality-based simulators has been widely adopted for training in various laparoscopic [10–14], robotic [15–17], and endoscopic procedures [18, 19]. However, the application of this technology to open surgery remains significantly limited [20]. We are in the process of developing a virtual reality-based trainer specifically for colorectal surgery (VCOST) to facilitate assessment and training in selected open colorectal surgical tasks. In this manuscript, we will explore the development of task-specific objective metrics through expert consensus for the automated assessment of perineal proctectomy using the Altemeier procedure. This study aims to assess the efficacy of these task-specific metrics in differentiating between novice and expert surgeons.

Materials and methods

Approval for this study was granted by the UT Southwestern institutional review board (single IRB #STU-2021-0202), and it was conducted in three distinct phases. During phase I, experts were interviewed to establish task-specific metrics for assessing performance of the Altemeier procedure for rectal prolapse. In phase II, expert consensus on the developed metrics was obtained through a survey, and in phase III, an inanimate model study was conducted to assess the validity evidence supporting the task-specific metrics.

Development of task-specific metrics for altemeier procedure for rectal prolapse

In phase I of this IRB-approved study, we performed a hierarchical task analysis for the Altemeier procedure. Such analysis is a widely recognized method that dissects surgical procedures into tasks, subtasks, and motion-end effectors [21, 22]. This involved in-depth interviews with expert colorectal surgeons affiliated with Baylor University Medical Center. Supplementary procedural insights were gathered from surgery textbooks and instructional workshop videos. The primary tasks and subtasks of the procedure were identified and refined under the guidance of an expert colorectal surgeon. Through this iterative process, a total of 31 metrics, grouped into task-specific (25) and general (6) metrics were developed. A Likert scale utilizing a 5-point system was employed to evaluate each task-specific metric, where a score of 5 denoted complete correctness and a score of 0 indicated complete incorrectness. Intermediate points between 5 and 0 were considered as indicating sub-optimal performances.

In phase II of the study, consensus on the importance of the developed metrics was obtained from practicing colorectal surgeons with varying years of experience (< 5, 5–15, and > 15 after completion of their training) using an online survey administered using the SurveyMonkey platform. Weighted averages of the ratings of each metric item on a 5-point Likert scale, with 1 being the least important and 5 the most important, were used to assess the importance of the developed metrics.

Validation of the perineal proctectomy task-specific metrics

During phase III, we scrutinized the validity of the task-specific metrics through a study conducted at UT Southwestern in its Artificial Intelligence and Medical Simulation Lab, utilizing a custom-made rectal prolapse model. To assess the validity, Messick's unitary framework was employed [23]. Our focus was on evaluating validity evidence across various domains, including content alignment, response process, internal structure, and the relationship to other variables.

Rectal prolapse simulator design

We developed an inanimate rectal prolapse model using biotissue bowel to create the rectal prolapse through the anus of a 3D-printed anatomic model, secured on a 3D-printed stand (Fig. 1a). A hemostat was used to make an indentation to represent the dentate line (Fig. 1b). A pair of standard needle drivers, forceps, hemostats, and scissors were available for

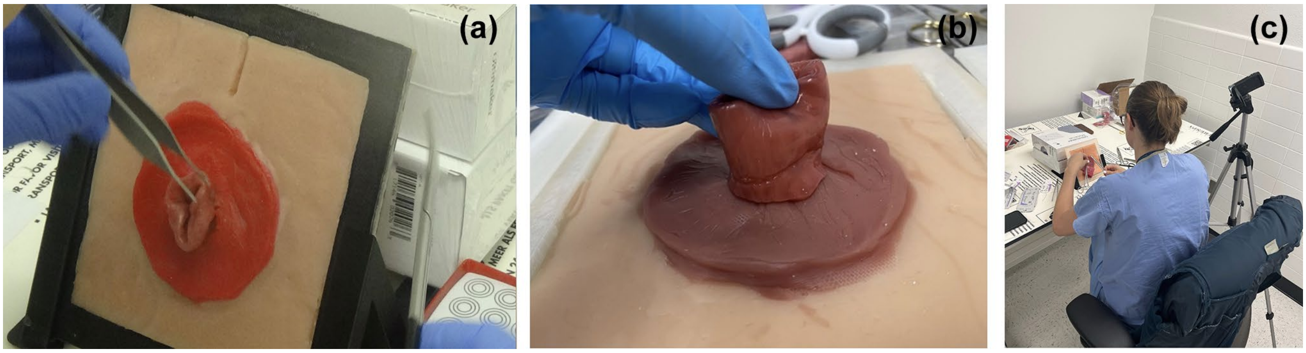


Fig. 1 **a** and **b** Inanimate rectal prolapse model and **c** participant performing Altemeier procedure on the model

participant use to perform the Altemeier procedure. In addition, 2–0 vicryl sutures were provided for suturing. A camera mounted on a tripod was positioned to exclusively capture the participants' hands during the procedure, ensuring anonymity for subsequent video-based assessments (Fig. 1c).

Study design and procedure

The study was performed at UT Southwestern in the Artificial Intelligence and Medical Simulation Lab with general surgery residents and colorectal attendings as participants. The participants recruited were separated into two groups by level of expertise: novice (general surgery junior PGY1-3 residents) and expert (senior general surgery PGY4-5 residents and colorectal faculty surgeons). After completing informed consent, each participant filled out a pre-survey to report demographic information, previous clinical exposure, and simulator experience.

No operative or technical guidance was given regarding the procedure since the steps of the procedure were part of the task-specific assessment, but we did provide general instructions explaining the study's aims. The participants were then asked to perform an unassisted perineal proctectomy on our simulator. There was a 1-h time limit to complete up to two attempts. Video recordings were limited to the participants' hands with the instruments actively utilized, to the biotissue model, and to a card revealing their random subject identification number to keep each participant completely de-identified during the procedure.

After completing the simulated procedure, participants filled out a post-survey to evaluate the quality of the simulator on a 5-point Likert scale questionnaire. The survey covered 5 categories that included the visual appearance of the simulation, the quality of models and textures, the realism of the simulator interface, how closely the task mirrored the actual surgical procedure, and the simulator's overall effectiveness in teaching the Altemeier procedure. Participants also completed a 10-point NASA Task Load Index post-survey to assess their perceived cognitive task load. A Mann–Whitney U test was used to assess the difference between the groups.

We had two blinded qualified raters perform video-based assessment using the task-specific metrics (Table 1) as well as global rating metrics (Table 2). The latter was derived from the Objective Structured Assessment of Technical Skills (OSATS), which was previously validated for the assessment of technical skills in colorectal surgery [24]. The raters initially assessed the performance of participants in 5 videos, compared their ratings, and resolved any discrepancies. Subsequently, they evaluated the quality of another 5 videos before completing the grading process for the remaining videos. An intraclass correlation coefficient (ICC) based on mean rating ($k = 2$) as well as absolute agreement and a two-way mixed-effects model were used to assess the agreement between the two raters to establish interrater reliability (IRR). An ICC value of 0.75 to 0.9 was considered good and above 0.9 as excellent for IRR [19].

Data analysis

In phase II, which involved consensus development of task-specific metrics, descriptive statistics were used to analyze the data from the survey, and a weighted average of importance scores from the survey was calculated for all the metric items. The responses from colorectal surgeons within the three experience levels (< 5 years, 5–15 years, and > 15 years) were analyzed using a nonparametric Kruskal–Wallis test. Post-hoc analysis of significant results was performed using the Wilcoxon rank-sum test. To control the false discovery rate due to multiple comparisons, a Benjamini and Hochberg correction [25] was used to calculate the adjusted P values with significance set at 0.05.

In phase III, for both global and task-specific scores, a total score was derived by summing all the individual domains. To assess the correlation between the global and task-specific metrics in the evaluation of technical skills, we performed a Spearman's rank correlation test.

Table 1 Task-specific metrics used for assessment of performance

	Metric	Score (0, 3 or 5 points)		
		5	3	0
1	Initial incision	Circumferential cutting 1 cm proximal to the dentate line	Circumferential cutting > 1 cm or < 1 cm to the dentate line	No cutting
2	Force on tissue	Accurate force	Struggle to remove	Tears rectum
3	Filling the intervening spaces	Stay sutures are used to complete anastomosis with good apposition of colon and anal mucosa	–	Incomplete anastomosis
4	Closure of the sutures	Successful closure	–	Inadequate closure
5	Rectal perforation during the dissection	No perforation	–	Perforation
6	Tool handling	Economy of moves	Some unnecessary moves	Unnecessary moves
7	Motion	Deficient knowledge	Knew all important steps of procedure	Demonstrated familiarity with all aspects of procedure
8	Suture handling	Equidistant placement of full-thickness sutures	Poor placement of sutures	Inadequate suture handling
9	Task execution order	Completion of tasks executed in order	–	Completion of tasks not executed in order
10	Task completion	Tasks are done	Tasks are partially done	Tasks are not done

Items 3,4,5, and 9 could only be scored as 0 or 5

A Wilcoxon rank-sum test was used to evaluate the differences in the performance of novice and expert participants.

Sample size

An a priori power analysis was conducted using G*software [26] to test the difference in performance between the three groups, with $\alpha = 0.05$, an effect size of $f = 0.7$ and a power of $\beta = 0.8$. The analysis showed that a total of 20 subjects equally distributed in two groups was needed to achieve the necessary power.

Results

Phase I: hierarchical task analysis results

The hierarchical task analysis was carried out to delineate the essential steps of the Altemeier procedure and organize them in a hierarchical sequence. A task tree (Fig. 2) was formulated to establish the optimal sequence of executing steps and sub-steps, showing their linear progression. Utilizing hierarchical task analysis, performance metrics were developed. The analysis revealed four distinct primary steps in the Altemeier procedure: (1) preparation, (2) circumferential cutting, (3) dissection, and (4) resection. The details of each of these steps are shown in Fig. 2.

Phase II: expert consensus survey results

A total of 23 colorectal surgeons with varying years of experience (< 5 years = 1 surgeon; 5–15 = 6; > 15 = 16) participated in this study. Most participants (87%) reported that using a retractor is their preferred method of performing the procedure. When asked how close to the dentate line they begin circumferential cutting, 52% reported 1 cm proximal, 39% > 1 cm, and 9% < 1 cm. The weighted average scores for the 25 task-specific and 6 general metrics are shown in Table 3, and answers are grouped by experience in Fig. 3a, b, respectively. Between the groups, agreement was found in the ratings of importance for all metric items ($P > 0.05$).

Most of the surgeons (84%) preferred to use a retractor when performing a perineal proctectomy, 12% did not use a retractor, and 4% did not have a preference. When asked about the importance of time in the execution of the task, 52% indicated it was somewhat important, 20% didn't think it was important, 8% indicated it was very important, and 20% remained neutral. When asked how much time is adequate to perform this task, 84% indicated more than 30 min but less than an hour is an adequate time to perform the perineal proctectomy procedure; however, 16% reported this time depended on several factors. In regard to the circumferential incision required to perform the procedure, nearly half of the surgeons (48%) reported they begin at 1 cm, 40% reported more than 1 cm, and 12% reported less than 1 cm proximal to the dentate line. Regarding how important it is to place the patient in a prone-flexed position before starting

Table 2 Global metrics used for assessment of performance

	Domain of surgical performance	Score (5-point Likert scale)				
		1	2	3	4	5
1	Respect for tissue	Frequently used unnecessary force on tissue or caused damage	–	Careful handling of tissue but occasionally caused inadvertent damage using more force than needed	–	Consistent handling of tissue, minimizes damage through appropriate use of instruments and appropriate force
2	Time and motion	Many unnecessary moves	–	Efficient time/motion but some unnecessary moves	–	Clear economy of movement and maximum efficiency
3	Instrument handling	Repeatedly made tentative or awkward moves with instruments	–	Competent use of instruments but occasionally appeared stiff or awkward	–	Fluid moves with instruments and no awkwardness
4	Flow of operation	Frequently stopped procedure and seemed unsure of next move	–	Demonstrated some forward planning with reasonable progression of procedures	–	Obviously planned course of procedure with effortless flow from one move to the next
5	Knowledge of specific procedure	Deficient knowledge	–	Knew all important steps of procedure	–	Demonstrated familiarity with all aspects of procedure
6	Overall performance	Very poor	–	Competent	–	Expert level

A score of 2 or 4 indicates performance that was intermediate to the performance shown in adjacent cells

the procedure, 30.4% did not think this was important at all, 30.4% thought it was somewhat important, and 17.39% thought it was very important.

Vicryl was the most preferred suture material to close the inner wall, followed by polydioxanone and chromic (preferred by 78.26%, 17.39%, and 4.35% of surgeons, respectively). The outer wall closure was most commonly done using vicryl (73.91%), followed equally by silk and polydioxanone (each 13.04%). The preferred method for closing the cut end of the rectum was sutures (58.33%), followed by an Allis clamp (25%) and no preference (16.67%). When placing sutures in the anal cuff, 83.33% placed four-quadrant stay sutures through mucosa and internal sphincter, while 12.5% did not place them. In regard to perineal dissection of the posterior precoccygeal plane, 58.33% started their dissection posteriorly and then proceeded laterally and anteriorly, and 37.5% started anteriorly and then proceeded posteriorly. Regarding the importance of resecting the left colon at the level of the anal canal with moderate tension on the colon, 41.67% were neutral, 29.17% reported it was somewhat important, 20.83% reported it was very important, and 8.33% reported it was not important.

Phase III: metrics validation study results

Pre-survey results

Demographics A total of 22 participants were recruited to complete an Altemeier procedure using our inanimate rectal prolapse model (Fig. 1). Of those, 45% ($n=10$) were female, 36% ($n=8$) identified as being white race, and 73% ($n=16$) wore corrective lenses. Participants were grouped into novice ($n=15$) and expert ($n=7$). Table 4 shows the demographic characteristics of surgical residents and faculty who participated in this study.

Prior experience Among the expert attendings, 1 reported having performed 15–20 perineal proctectomies, while 1 reported performing 10–15, and 1 reported performing 5–10. The rest of the expert group ($n=4$; 57%) consisted of PGY4-5 residents, and they reported performing 0–5. All of the novice group ($n=14$; 93%) reported observing and performing 0–5 perineal proctectomies, except for 1 who reported observing 10–15 and performing 5–10 of the operations. Overall, 73% ($n=16$) of participants self-reported having subjective experience of prior exposure to a either a robotic (DaVinci robot), laparoscopic (FLS), and/or open ($n=5$; 23%) simulation trainer. These self-reported accounts indicate that a subset of the participants had previous hands-on engagement or familiarity with the technology being assessed. Additionally, 30.3% ($n=12$; 54%) reported having gaming experience, with almost half of them ($n=4$; 18%) playing at least 1–5 h a week.

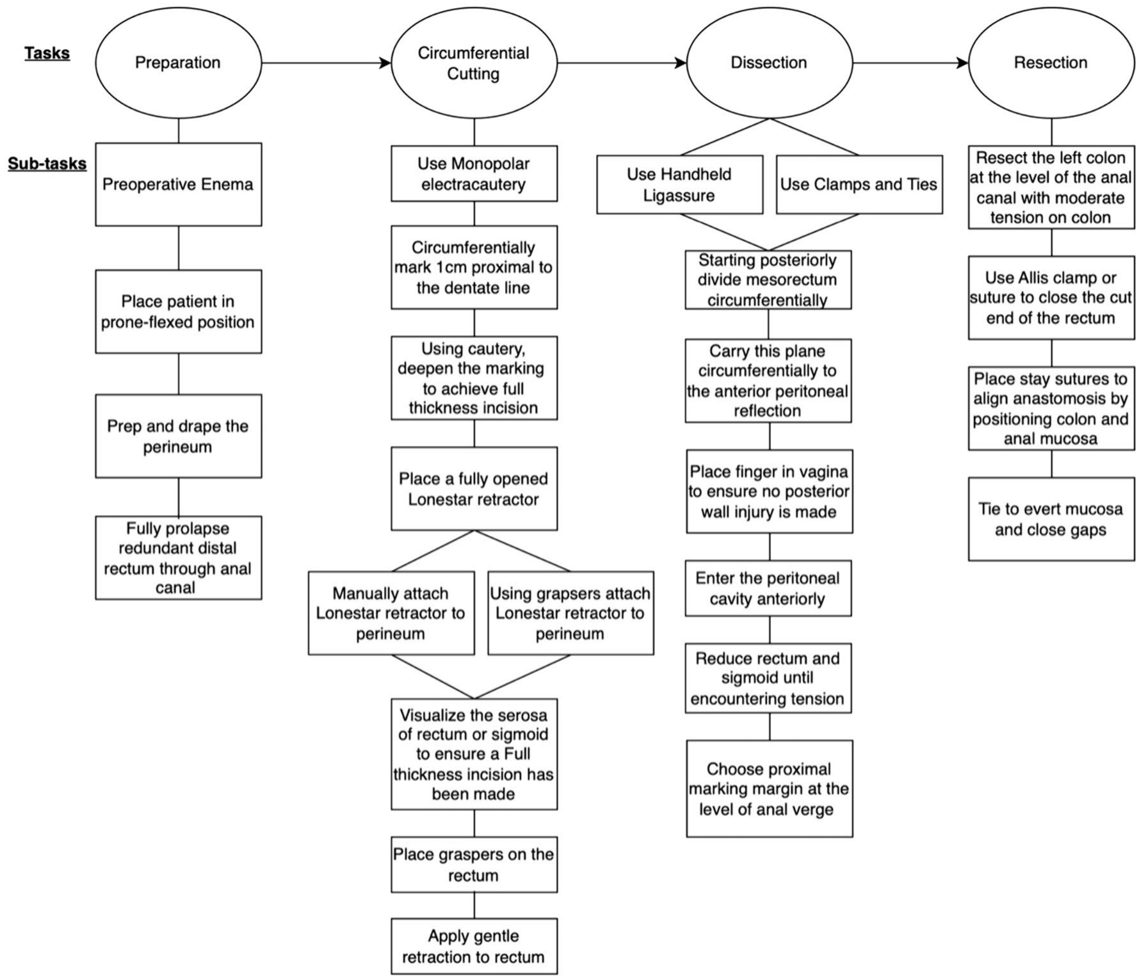


Fig. 2 Hierarchical task analysis results of the Altemeier procedure for rectal prolapse

Post-survey results

Following the completion of the perineal proctectomy simulation, a post-survey was administered where participants evaluated their experience using a Likert scale. The scale ranged from 1 (not realistic) to 5 (very realistic). The assessment covered five categories: realism of the model's anatomy, texture realism of the synthetic model, realism of the simulator interface (including instruments and display), overall realism compared to the actual surgical task, and the perceived usefulness of the simulator in learning open perineal proctectomy surgical skills. Table 5 displays the survey results, indicating the assessed levels of realism and usefulness of the perineal proctectomy simulation model.

One participant rated the degree of realism of the anatomy as 5/5, and only 6 (27%) rated the anatomy as 4/5 (Table 5). The majority of the group rated it as 3/5 or below ($n=15$; 68%). When assessing the degree of realism of the model and its texture, a third of the participants rated it at or above 4/5 ($n=9$; 41%), while the rest of the group rated it as 3/5 or below. When evaluating the overall realism of the simulator interface, almost half of the group ($n=9$; 41%), rated the task as 4/5 or higher. When evaluating the overall realism of the simulation compared to the actual surgical task, 91% ($n=20$) of the participants rated the task as 3/5 or higher, while about a third of the group ($n=9$; 41%) rated it as 4/5 or higher. Finally, when assessing the perceived usefulness of the simulator in learning how to perform a perineal proctectomy, the majority of participants ($n=13$; 59%) rated it as 4/5 or 5/5 (Table 5).

Cognitive task load results

After each participant completed the perineal proctectomy on the inanimate model, they were asked to complete a 10-point NASA Task Load Index post-survey to assess their perceived cognitive workload. Comparison of median scores showed that experts had significantly less workload compared to novices (index total score: 21 vs. 30; $P=0.01$). In individual domains (Fig. 4), experts

Table 3 Weighted average of each metric for perineal proctectomy

Metric number	Metric	Weighted average
M1	Position of patient	2.94
M2	Placing Lone Star retractor	3.69
M3	Prolapsing distal rectum	4.56
M4	Instrument for incision of distal rectum	3.38
M5	Initial incision	4.13
M6	Transection of layers of rectum	4.31
M7	Grasping force on prolapsed tissue	4.5
M8	Circumferential incision in the rectum	4.75
M9	Placing sutures	4.06
M10	Selecting sutures	3.25
M11	Dissection in the areolar tissue plane posteriorly	4.44
M12	Enter cul de sac between rectum and vagina	4.75
M13	Tagging the apex of the cul de sac anteriorly	2.73
M14	Incise the peritoneum of the pelvis	3.75
M15	Release the posterior mesorectum	4.63
M16	Release the descending colon to reach the pelvis	3.69
M17	Resection	4.4
M18	Closing the cut end of the rectum to prevent contamination	2.69
M19	Filling the intervening spaces	4.5
M20	Closure of suture	4.6
M21	Removing the tools	4.19
M22	Rectal perforation during the dissection	4.38
M23	Ligation of mesorectal vessels at rectosigmoid junction	4.38
M24	Control of bleeding	4.81
M25	Bleeding intervention	4.69
M26	Tool handling	4.44
M27	Economy of motion	4.25
M28	Suture handling	4.31
M29	Knowledge of instrument and procedure	4.19
M30	Tasks execution order	4.38
M31	Repetition time	3.75

had significantly less physical demand (2 vs. 4; $P = 0.01$) and temporal demand (1 vs. 5; $P = 0.002$). Though not statistically significant, experts also expressed less mental demand (3 vs. 6; $P = 0.07$) and frustration (3 vs. 5; $P = 0.25$) compared to novices. There were no significant differences in performance (5 vs. 5; $P = 0.72$) or effort (5 vs. 5; $P = 0.47$).

Reliability analysis

The IRR for the two raters using ICC was high and showed that both global (ICC = 0.97; $P < 0.001$) and task-specific (ICC = 0.95; $P < 0.001$) metrics had achieved the minimum requirement for good reliability, thus establishing response process validity.

Analysis of metrics

Table 6 depicts the multidimensional metrics and descriptive statistics used to evaluate the performance of participants.

Global metrics

Experts performed better than novices (Fig. 5a) when evaluated using the global metric scores (12 vs. 24; $W = 19.5$; $P = 0.02$).

Task-specific metrics

Experts performed better than novices (Fig. 5b) as well when evaluated using the task-specific metrics (39.5 vs. 70; $W = 23.5$; $P = 0.04$).

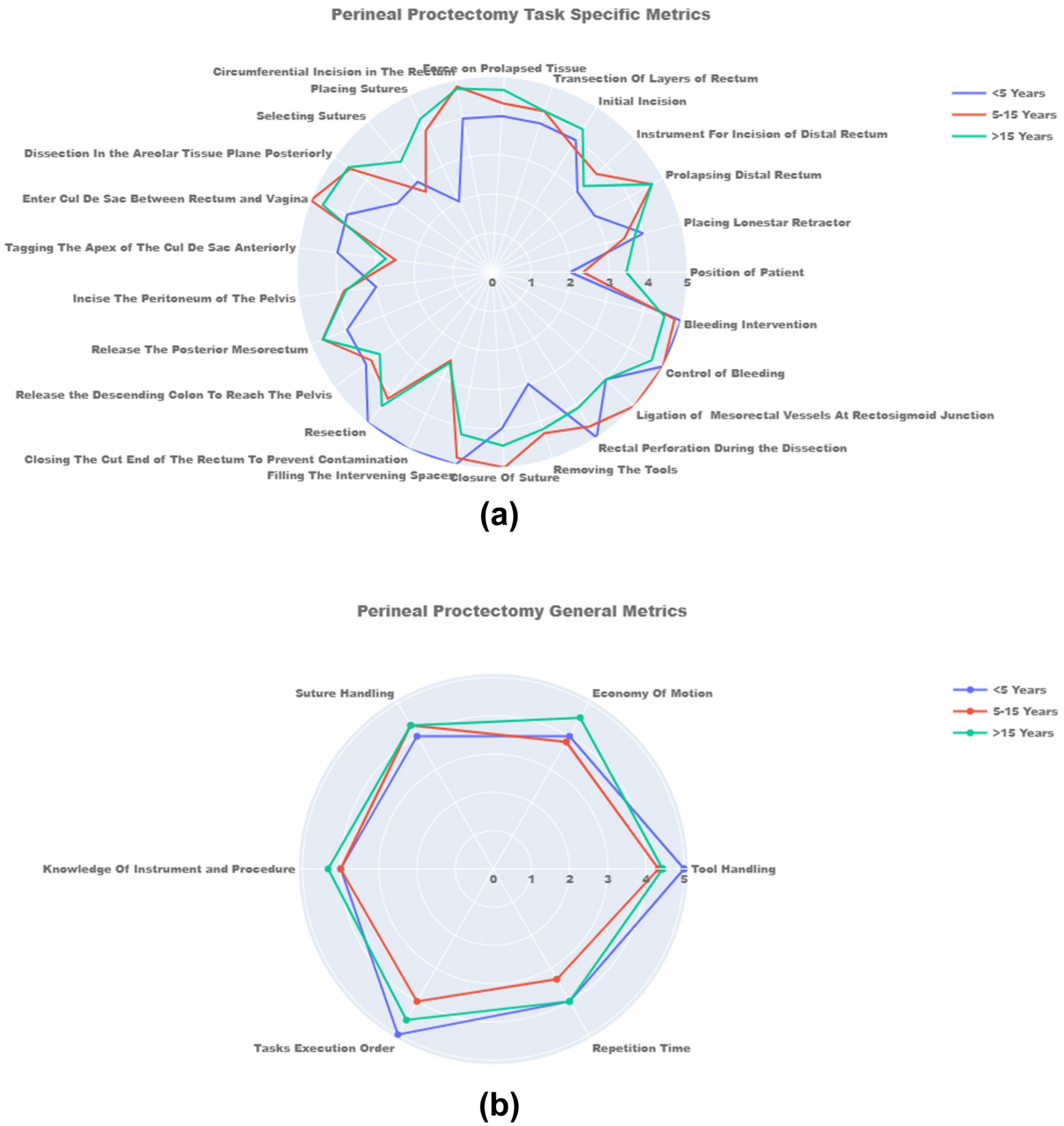


Fig. 3 Weighted average of ratings of surgeons with < 5, 5–15, or > 15 years of experience regarding perineal proctectomy. **a** Task-specific metrics and **b** general metrics

Correlation between global and task-specific metrics

A correlation analysis using the Spearman rank correlation coefficient showed high correlation ($R = 0.9$; $P < 0.001$), indicating a strong positive association between global and task-specific metrics (Fig. 6).

Discussion

Our results demonstrate the effectiveness of our task-specific metrics in differentiating between novice and expert operators performing a perineal proctectomy using an inanimate rectal prolapse model. The strong positive correlation observed between global and task-specific scores emphasizes

Table 4 Demographics for the participants

	Novice	Expert	Total
Number of participants	15	7	22
Sex, female, *n* (%)	6 (40)	4 (57)	10 (45)
Age, mean, years	30	33	32
Race, white, *n* (%)	5 (33)	3 (43)	8 (36)
Ethnicity, Hispanic, *n* (%)	3 (20)	0	5 (15)
Corrective lenses, yes, *n* (%)	10 (67)	6 (86)	16 (73)

Table 6 Median and interquartile range (IQR) of metrics used for the assessment of performance

Metric	Group	Median [IQR]	Wilcoxon rank-sum test
Total global score	Novice	12 [4.75]	$W=10.5$
	Expert	24 [9.5]	$P=0.02$
Task-specific score	Novice	39.5 [17.25]	$W=23.5$
	Expert	70 [17]	$P=0.04$

their mutual influence, highlighting the necessity of considering both aspects in evaluating trainee performance. Noteworthy between-group distinctions were evident for both global and task-specific scores, supported by a high level of interrater reliability, reinforcing the robustness of our findings. This study also reaffirms the viability of employing a trained observer to assess task performance through video-based assessment to determine an operator's proficiency in the specific task at hand. Consequently, it becomes feasible to design tasks within a simulation model that effectively differentiates and assesses operator performance in a training program context.

The ability to determine proficiency in this specific procedure is important because it is often reserved for high-risk patient populations. Mickulicz initially introduced perineal proctectomy (rectosigmoidectomy) as a treatment for full-thickness rectal prolapse in 1889; however, it took more than 8 decades for this approach to gain significant recognition in the United States, notably following the report by Altemeier in 1971 [27]. Subsequently, surgeons have commonly

Table 5 Survey completed after performing the perineal proctectomy simulation on the inanimate model

Score from 1 (not realistic) to 5 (very realistic)	1/5	2/5	3/5	4/5	5/5
Realism of the anatomy of the model, *n* (%)	0 (0)	4 (18)	11 (50)	6 (27)	1 (5)
Realism of the synthetic model (texture), *n* (%)	1 (5)	5 (23)	7 (32)	7 (32)	2 (9)
Realism of the simulator interface (instrument, display), *n* (%)	1 (5)	0 (0)	10 (45)	5 (23)	6 (27)
Overall realism of the task compared to the actual surgical task, *n* (%)	2 (9)	1 (5)	10 (45)	6 (27)	3 (14)
Overall usefulness of the simulator in learning laparoscopic hiatal hernia skills, *n* (%)	1 (5)	0 (0)	8 (24)	6 (27)	7 (32)

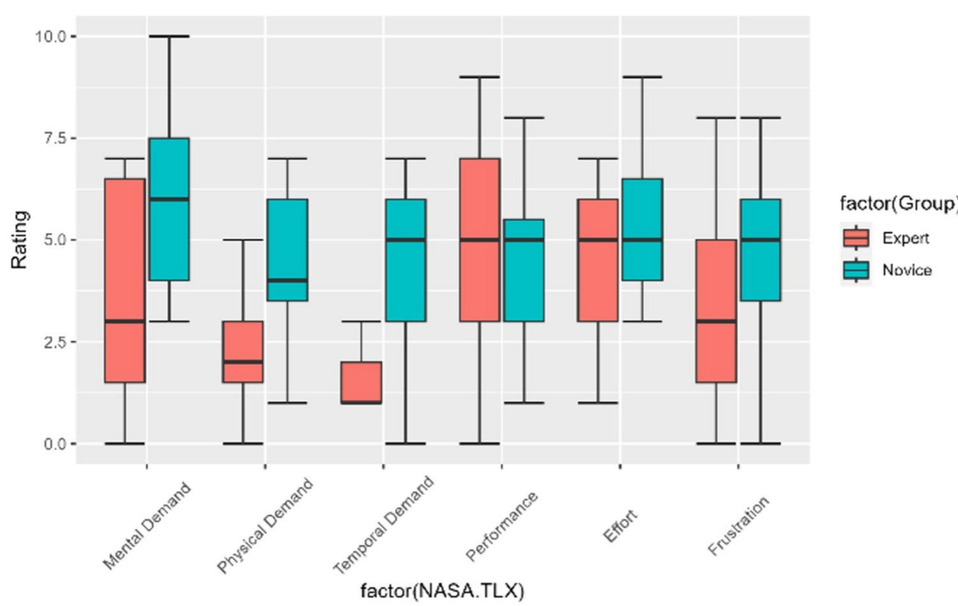


Fig. 4 Box plot of NASA Task Load Index ratings for both groups with the X axis listing the index's factor and the Y axis denoting participant ratings

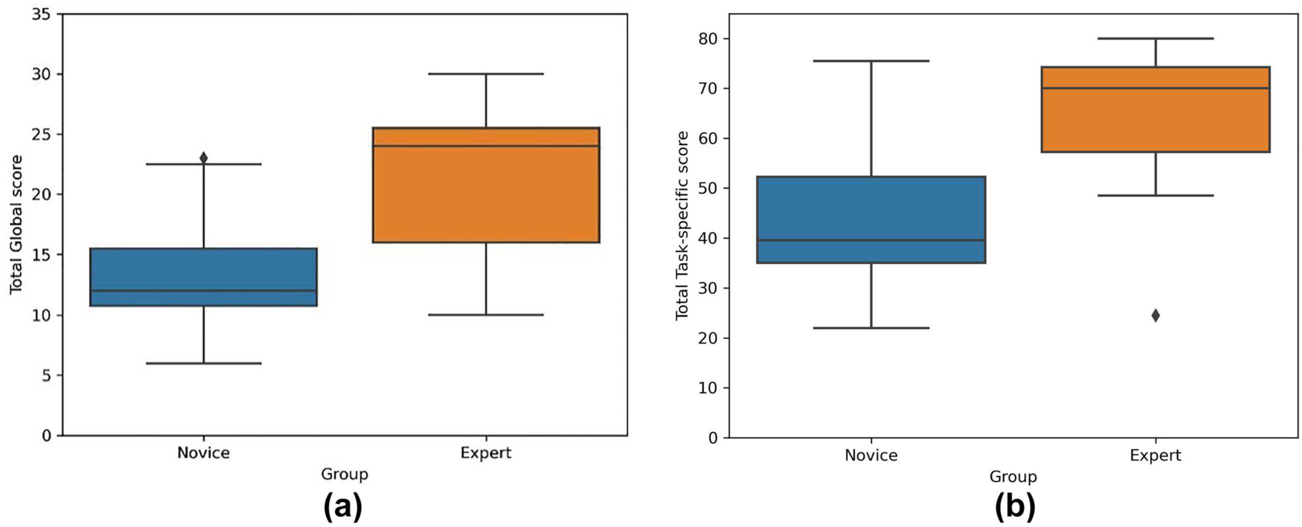
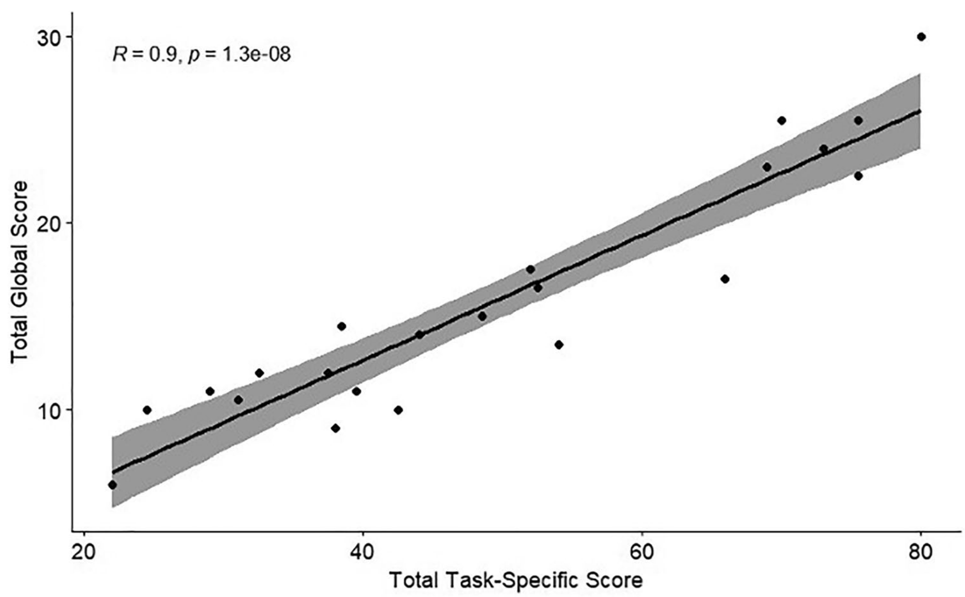


Fig. 5 **a** Global and **b** task-specific total scores of performance

Fig. 6 Correlation between global and task-specific scores



reserved perineal proctectomy, along with other perineal-only methods, for elderly, institutionalized, or high-comorbid patients considered too high-risk for transabdominal procedures [28]. Furthermore, this approach is frequently the preferred surgical method for patients presenting with acute incarceration and strangulation of the prolapsed rectum [29]. The approach has persisted despite the additional options of various laparoscopic and robotic rectopexy repairs and resections [30, 31]. Despite the array of treatment alternatives for rectal prolapse and the considerable absence of well-conducted randomized studies, both recent Cochrane reviews on the subject were unable to affirm or challenge the perceived superiority of transabdominal repairs for this condition [32–34]. Noteworthy is the observation that perineal procedures continue to account for approximately 50% to 60% of the total procedures performed for rectal prolapse throughout the United States [35]. Recurrence stands out as the most prevalent complication of this procedure, affecting approximately 23% [36]. This is of great concern, considering that each hospital admission for this frail population contributes to its increased risk of morbidity and even mortality [37–41]. Therefore, despite the growth of robotic and laparoscopic surgery, this is still an incredibly important open procedure to be not only familiar with but proficient in performing. The data from this study are contributing to the development of a virtual reality perineal proctectomy simulator that could fill a potential clinical exposure gap, as all but 1 participant of the novice group of our study reported observing and performing 0–5 perineal proctectomies.

The value of the simulator will be the built-in task-specific and global metrics that provide the discriminate validity for operators. These metrics follow the primary steps from the hierarchical task analysis survey, which include (1) preparation, (2) circumferential cutting, (3) dissection, and (4) resection. An analysis of results showed that our task-specific metrics distinguished performance between our novice and experienced groups. Correlation analysis showed a high correlation to the global metric item.

While our inanimate simulation received a lower score in terms of “realism,” this outcome was expected, considering the absence of additional anatomical components present in live patients, such as the mesorectum, taenia coli, connective tissue, and blood vessels. The incorporation of these elements into the virtual reality simulator is planned for future iterations. Nevertheless, it is reassuring that our perineal proctectomy simulation closely mirrors this procedure in real life, which includes the associated task-specific metrics. All participants except 2 rated the simulation's realism compared to the actual surgical task as being above 3/5, and the majority gave the simulator a rating above 4/5 for perceived usefulness in learning how to perform a perineal proctectomy.

Despite the favorable feedback received, this study has several acknowledged limitations. Primarily, the sample size was small, and participants were unevenly distributed among the groups. This uneven distribution is attributed to the limited number of actively practicing colorectal surgeons and the presence of only one colorectal fellow in our institution, who was engaged in conducting the study and consequently was unable to participate. Additionally, some participating attending surgeons had minimal recent exposure to open perineal proctectomy, potentially impacting the overall expertise represented in the study. Recruiting senior residents posed a challenge due to increasing operative and clinical responsibilities with each postgraduate year. Moreover, there may have been reluctance, reduced study recruitment, and concerns about skill judgment among residents, given that their co-resident conducted the study and graded the videos, even though the videos were blinded for evaluation. These limitations could have affected the generalizability and robustness of the study's findings.

In summary, our findings provide evidence of discriminant validity for both task-specific and global metrics in an innovative inanimate rectal prolapse model. This study advocates for the continued refinement of this task and further validation studies with more participants to establish the feasibility of this model to facilitate proficiency-based training. The creation of a virtual reality simulator for perineal proctectomy has the potential to transform surgical training and skill acquisition significantly. Our next objective involves integrating these validated metrics into our virtual reality simulator for automated assessment.

Acknowledgements We thank Dave Primm of the UT Southwestern Department of Surgery for help in editing this manuscript. We would also like to acknowledge the support from NIH/NIBIB R01: EB025241.

Funding This work was funded by NIH/NIBIB R01: EB025241.

Declarations

Disclosures Alexis Desir, Poya Pourghaderi, Shruti Doga Demirel, Javier Salgado Pogacnik, Suvranu De, James Fleshman, Ganesh Sankaranarayanan have no conflicts of interest or financial ties to disclose.

References

1. Birkmeyer JD et al (2013) Surgical skill and complication rates after bariatric surgery. N Engl J Med 369(15):1434–1442
2. Ray JJ et al (2016) Association between American Board of Surgery In-Training Examination scores and resident performance. JAMA Surg 151(1):26–31
3. Fernandez GL et al (2012) Boot cAMP: educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. J Surg Educ 69(2):242–248
4. Abdelsattar JM et al (2018) Correlation of objective assessment data with general surgery resident in-training evaluation reports and operative volumes. J Surg Educ 75(6):1430–1436
5. Feldman LS et al (2004) Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg 198(1):105–110
6. De Montbrun SL et al (2013) A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the Colorectal Objective Structured Assessment of Technical Skill (COSATS). Ann Surg 258(6):1001–1006
7. Demirel D et al (2016) A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer simulator. Am J Surg 212(3):475–484
8. Sankaranarayanan G et al (2023) Objective metrics for hand-sewn bowel anastomoses can differentiate novice from expert surgeons. Surg Endosc 37(2):1282–1292
9. Sankaranarayanan G et al (2022) Development and validation of task-specific metrics for the assessment of linear stapler-based small bowel anastomosis. J Am Coll Surg 235(6):881–893
10. Satava RM (1993) Virtual reality surgical simulator. In: Surgical endoscopy. Springer, New York, pp 203–205
11. Van Dongen K et al (2007) Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 21:1413–1417
12. Arikatla VS et al (2014) Towards virtual FLS: development of a peg transfer simulator: development of VBLaST-PT© simulator. Int J Med Robot Comput Assist Surg 10:344–355
13. Nemani A et al (2018) Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator. Surg Endosc 32:1265–1272
14. Wynn G, Lykoudis P, Berlingieri P (2018) Development and implementation of a virtual reality laparoscopic colorectal training curriculum. Am J Surg 216(3):610–617

15. Lerner MA et al (2010) Does training on a virtual reality robotic simulator improve performance on the da Vinci® surgical system? J Endourol 24(3):467–472
16. Sethi AS et al (2009) Validation of a novel virtual reality robotic simulator. J Endourol 23(3):503–508
17. Perrenot C et al (2012) The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. Surg Endosc 26:2587–2593
18. Dorozhkin D et al (2016) Face and content validation of a virtual translumenal endoscopic surgery trainer (VTEST™). Surg Endosc 30:5529–5536
19. Ritter EM et al (2003) Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 7:871–878
20. Sankaranarayanan G et al (2016) Face and construct validation of a next generation virtual reality (Gen2-VR(©)) surgical simulator. Surg Endosc 30:979–985
21. MacKenzie L, Ibbotson JA, Cao CGL, Lomax AJ (2001) Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Minim Invasive Ther Allied Technol 10(3):121–127
22. Cao CG et al (1999) Hierarchical decomposition of laparoscopic procedures. Stud Health Technol Inform 62:83–89
23. Messick S (1995) Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 50(9):741
24. Martin J et al (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84(2):273–278
25. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57(1):289–300
26. Faul F et al (2007) G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39(2):175–191
27. Altemeier W et al (1971) Nineteen years' experience with the one-stage perineal repair of rectal prolapse. Ann Surg 173(6):993
28. Brown A et al (2004) Strategy for selection of type of operation for rectal prolapse based on clinical criteria. Dis Colon Rectum 47:103–107
29. Sipahi M et al (2016) Perineal rectosigmoidectomy for incarcerated rectal prolapse (Altemeier's procedure). Turk J Surg 32(3):217
30. Purkayastha S et al (2005) A comparison of open vs. laparoscopic abdominal rectopexy for full-thickness rectal prolapse: a meta-analysis. Dis Colon Rectum 48:1930–1940
31. Kariv Y et al (2006) Long-term outcome after laparoscopic and open surgery for rectal prolapse: a case–control study. Surg Endosc Int Tech 20:35–42
32. Tou S et al (2008) Surgery for complete rectal prolapse in adults. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD001758.pub2
33. Bachoo P, Brazzelli M, Grant A (2000) Surgery for complete rectal prolapse in adults. Cochrane Database Syst Rev 20(2):CD001758
34. Young MT et al (2015) Surgical treatments for rectal prolapse: how does a perineal approach compare in the laparoscopic era? Surg Endosc 29:607–613
35. Schoetz DJ Jr (2006) Evolving practice patterns in colon and rectal surgery. J Am Coll Surg 203(3):322–327
36. Tiengtianthum R et al (2014) Clinical outcomes of perineal proctectomy among patients of advanced age. Dis Colon Rectum 57(11):1298–1303
37. Franchi C et al (2013) Risk factors for hospital readmission of elderly patients. Eur J Intern Med 24(1):45–51
38. Berian JR et al (2016) Association of loss of independence with readmission and death after discharge in older patients after surgical procedures. JAMA Surg 151(9):e161689–e161689
39. Tsai TC, Orav EJ, Jha AK (2015) Care fragmentation in the post-discharge period: surgical readmissions, distance of travel, and postoperative mortality. JAMA Surg 150(1):59–64
40. Rothenberg KA et al (2019) Association of frailty and postoperative complications with unplanned readmissions after elective outpatient surgery. JAMA Netw Open 2(5):e194330–e194330
41. Turina M et al (2013) Quantification of risk for early unplanned readmission after rectal resection: a single-center study. J Am Coll Surg 217(2):200–208