Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Surgical video and body composition do not yet add to conventional predictors of complications after robotic distal gastrectomy

Sung Hyun Park
Yonsei University
Hannah Heejung Lee
Ludwig-Maximilian-University Hospital
Haemin Yong
Yonsei University
Kyungri Park
Hanyang University
Woohwan Jung
Korea University
Sang Won Lee
Yonsei University
Jawon Hwang
Yonsei University
Minah Cho
Yonsei University
Yoo Min Kim
Yonsei University
Si-Hak Lee
Pusan National University Yangsan Hospital
Sun-Hwi Hwang
Pusan National University Yangsan Hospital
Woo Jin Hyung
Yonsei University
Jae Hun Chung
Pusan National University Yangsan Hospital
Hyoung-Il Kim
cairus@yuhs.ac

Yonsei University

Research Article

Keywords: Gastrectomy, Robotic surgical procedures, Video recording, Body composition, Postoperative complications

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11745942/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Surgical video provides information on surgeon proficiency and operative workflow, and preoperative computed tomography on body composition. These measures have been associated with postoperative outcomes. Their added predictive value beyond conventional clinicopathological variables has not been tested with methods suited to the few complications in video cohorts. To explore this, we retrospectively analyzed 127 robotic distal gastrectomies by three surgeons at two centers in 2023–2024. The endpoint was any Clavien–Dindo grade II or higher complication within 30 days. Sixty-one candidate predictors formed four pre-specified blocks: conventional clinicopathological variables (12), video-rated KLASS-02-QC proficiency (3), computed tomography-derived body composition (4) and video-derived workflow (42). All 15 block combinations were compared by repeated nested cross-validation with elastic-net selection inside training folds, permutation testing and half-sampling. A complication occurred after 22 operations (17.3%). Combinations without conventional variables were not distinguishable from chance (nested area under the curve [AUC] 0.504–0.553) and retained no variable. Conventional variables alone gave the highest nested AUC (0.698, 95% range 0.636–0.743). Adding proficiency, body composition or workflow did not raise it (0.680, 0.684 and 0.647), nor did all three (0.626). Adding the workflow block raised the apparent AUC, estimated in the patients used to build the model, to 0.874, but not the nested AUC. Only American Society of Anesthesiologists physical status, sex and estimated blood loss had selection probabilities above 0.50. Conventional clinicopathological variables predicted 30-day complications, whereas the added predictive value of multimodal surgical video and body composition data remains to be clearly shown in larger cohorts.

Introduction

Minimally invasive distal gastrectomy, performed laparoscopically or robotically, is now a standard operation for gastric cancer [1–5]. Nevertheless, postoperative complications occur in up to one in five patients [4, 6]. These complications depend partly on how the operation is performed, yet this has been difficult to quantify. Perioperative risk assessment has relied on conventional clinicopathological variables, such as age, body mass index (BMI), comorbidity and tumor stage. These variables represent the operation only roughly, through operative time and estimated blood loss, and the patient's physique only through BMI.

Surgical video records the entire operation, providing rich but unstructured data on how it was performed [7]. In early approaches, experts rated the whole video for technical skill or for adequate performance of major steps, and such ratings were associated with complication rates [8, 9]. The Korean Laparoendoscopic Gastrointestinal Surgery Study Group 02 quality control (KLASS-02-QC) instrument applies this approach to D2 lymphadenectomy, and its score was associated with surgical outcomes in the KLASS-02 trial [10, 11]. These ratings, however, condense the operation into summary scores and do not describe how it unfolded over time. Recent studies have segmented operations into clinically meaningful phases to measure the duration and sequence of each step [12], and established observational frameworks also record disruptions to the surgical flow [13, 14]. In distal gastrectomy, such phase-based measures have been used to assess surgical complexity, efficiency and surgeon skill [15–18].

As a measure of the patient's physique, BMI does not distinguish fat from muscle. Preoperative computed tomography (CT) quantifies visceral fat, subcutaneous fat and skeletal muscle separately, and these measures have been more closely associated with complications after gastrectomy than BMI [19, 20]. A recent meta-analysis also found CT-defined sarcopenia to be independently associated with postoperative complications [21]. Surgical video and CT thus provide multimodal measures of both the operation and the patient. Although several of these measures are associated with postoperative outcomes and are increasingly easy to obtain, none is part of routine risk assessment.

Before these multimodal measures can be used for risk assessment, they must be shown to add predictive value beyond conventional clinicopathological variables. In video cohorts, this is difficult to demonstrate. A full workflow taxonomy produces several dozen variables per operation, whereas a cohort from one or two institutions typically includes about one hundred operations and only a few dozen complications. When candidate predictors outnumber events, a model fitted to the

whole cohort can appear highly discriminative [22, 23], and selecting variables before cross-validation inflates performance further [24, 25]. Associations reported under these conditions may reflect overfitting rather than true predictive value. To our knowledge, the added value of these measures has not been tested with methods suited to so few events.

In this study, we grouped conventional variables and these multimodal measures into four pre-specified blocks and compared every block combination with the conventional block alone (Fig. 1). Throughout validation, variable selection was confined to the training folds. The aim was to explore whether multimodal surgical video and body composition data add predictive value to conventional clinicopathological variables for 30-day complications after robotic distal gastrectomy. A secondary aim was to quantify how much apparent performance is inflated when these data are added.

Methods

Study design, patients and endpoint

We retrospectively analyzed robotic distal gastrectomies with D2 lymphadenectomy for which both a complete, unedited console recording and a complete clinical record were available. Exclusion criteria were conversion to open or laparoscopic surgery, a missing major step or an interrupted recording, image quality insufficient for phase-level annotation, and non-standard limited lymphadenectomy. The final cohort comprised 127 operations performed between January 2023 and December 2024 by three board-certified gastric surgeons at two tertiary referral centers (Table 1). Surgeon A was an experienced robotic gastric surgeon. Surgeons B and C were early in their independent robotic practice. Surgeon B had performed more than 100 laparoscopic gastrectomies but had no robotic training, whereas surgeon C had substantial exposure to robotic gastrectomy during residency and fellowship. All operations were performed with the da Vinci Xi system (Intuitive Surgical, Sunnyvale, CA, USA). The institutional review boards of both centers approved the study (4-2024-0769 and 55-2025-103) and waived the requirement for informed consent.

Table 1
Cohort characteristics by operating surgeon and by 30-day complication status

Characteristic	All (*n* = 127)	Surgeon A (*n* = 77)	Surgeon B (*n* = 40)	Surgeon C (*n* = 10)	p (surgeon)	No complication (*n* = 105)	Complication (*n* = 22)	p (outcome)
Block 1. Conventional clinicopathological variables								
Age (years)	61 (53–70)	61 (52–72)	59 (56–69)	64 (55–67)	0.848	60 (53–70)	64 (59–72)	0.178
Sex					0.628			0.007
Male	77 (60.6)	49 (63.6)	23 (57.5)	5 (50.0)		58 (55.2)	19 (86.4)	
Female	50 (39.4)	28 (36.4)	17 (42.5)	5 (50.0)		47 (44.8)	3 (13.6)	
Body mass index (kg/m²)	23.3 (21.1–25.5)	24.2 (21.3–25.7)	22.9 (20.2–24.9)	23.1 (22.7–25.2)	0.367	23.4 (21.3–25.3)	22.7 (20.4–25.7)	0.558
ASA physical status					0.070			0.002
I	26 (20.5)	13 (16.9)	12 (30.0)	1 (10.0)		26 (24.8)	0 (0.0)	
II	69 (54.3)	39 (50.6)	24 (60.0)	6 (60.0)		58 (55.2)	11 (50.0)	
III	32 (25.2)	25 (32.5)	4 (10.0)	3 (30.0)		21 (20.0)	11 (50.0)	
Previous abdominal surgery					0.105			0.760
Yes	21 (16.5)	12 (15.6)	5 (12.5)	4 (40.0)		17 (16.2)	4 (18.2)	
No	106 (83.5)	65 (84.4)	35 (87.5)	6 (60.0)		88 (83.8)	18 (81.8)	
Pathologic T category					0.046			0.381
T1	56 (44.1)	28 (36.4)	25 (62.5)	3 (30.0)		49 (46.7)	7 (31.8)	
T2	20 (15.7)	11 (14.3)	6 (15.0)	3 (30.0)		16 (15.2)	4 (18.2)	
T3	25 (19.7)	21 (27.3)	3 (7.5)	1 (10.0)		18 (17.1)	7 (31.8)	
T4	26 (20.5)	17 (22.1)	6 (15.0)	3 (30.0)		22 (21.0)	4 (18.2)	

Continuous variables are median (interquartile range), compared by the Kruskal–Wallis test across surgeons and by the Mann–Whitney U test by outcome. Categorical variables are *n* (%), compared by the chi-square or Fisher's exact test. Block 1 is the conventional predictor block defined in the Methods. Blocks 2 to 4 are summarized in Supplementary Table S1.

ASA American Society of Anesthesiologists

Characteristic	All (*n* = 127)	Surgeon A (*n* = 77)	Surgeon B (*n* = 40)	Surgeon C (*n* = 10)	p (surgeon)	No complication (*n* = 105)	Complication (*n* = 22)	p (outcome)
Block 1. Conventional clinicopathological variables								
Pathologic N category					0.053			0.247
N0	68 (53.5)	35 (45.5)	30 (75.0)	3 (30.0)		59 (56.2)	9 (40.9)	
N1	21 (16.5)	16 (20.8)	3 (7.5)	2 (20.0)		16 (15.2)	5 (22.7)	
N2	20 (15.7)	14 (18.2)	3 (7.5)	3 (30.0)		14 (13.3)	6 (27.3)	
N3	18 (14.2)	12 (15.6)	4 (10.0)	2 (20.0)		16 (15.2)	2 (9.1)	
Reconstruction					< 0.001			0.270
Billroth-I	28 (22.0)	15 (19.5)	9 (22.5)	4 (40.0)		26 (24.8)	2 (9.1)	
Billroth-II	64 (50.4)	54 (70.1)	6 (15.0)	4 (40.0)		49 (46.7)	15 (68.2)	
Billroth-II with Braun	21 (16.5)	7 (9.1)	14 (35.0)	0 (0.0)		18 (17.1)	3 (13.6)	
Roux-en-Y	14 (11.0)	1 (1.3)	11 (27.5)	2 (20.0)		12 (11.4)	2 (9.1)	
Operative time (min)	199 (164–269)	177 (152–214)	300 (268–330)	186 (161–204)	< 0.001	199 (166–270)	196 (154–226)	0.319
Estimated blood loss (mL)	50 (24–76)	50 (25–80)	35 (20–50)	68 (54–115)	0.022	45 (21–60)	70 (34–132)	0.027
Outcome								
Complication (Clavien–Dindo ≥ II)					0.238	—	—	—
Yes	22 (17.3)	15 (19.5)	4 (10.0)	3 (30.0)		—	—	
No	105 (82.7)	62 (80.5)	36 (90.0)	7 (70.0)		—	—	

Continuous variables are median (interquartile range), compared by the Kruskal–Wallis test across surgeons and by the Mann–Whitney U test by outcome. Categorical variables are *n* (%), compared by the chi-square or Fisher's exact test. Block 1 is the conventional predictor block defined in the Methods. Blocks 2 to 4 are summarized in Supplementary Table S1.

ASA American Society of Anesthesiologists

The endpoint was any postoperative complication of Clavien–Dindo grade II or higher within 30 days after surgery [26]. Complications were recorded prospectively and adjudicated from standardized postoperative records and, when necessary, by multidisciplinary review. This observed endpoint served as the reference standard for all estimates of discrimination and calibration. Twenty-two of the 127 operations met this definition, so every model in this study was fitted to 22 events.

Predictor blocks

Predictors were assigned to four pre-specified blocks, with conventional clinicopathological variables in block 1 and multimodal measures in blocks 2 to 4 (Fig. 1). The distributions of blocks 2 to 4 are shown in Fig. 2. All decisions described in this section were made without reference to the endpoint. Blocks 2 and 4 were derived from the same unedited console recordings, which two trained reviewers assessed independently while blinded to postoperative outcome. Discrepancies were resolved by a senior surgeon.

Table 2
Surgical phase taxonomy

Code	Phase	Description	Start criterion	End criterion
SAU01	Greater omentum (LN 4sb, 4d, LGEA)	Dissection of greater omentum	Energy device activation on omentum	Completion of left gastroepiploic lymphadenectomy
SAU02	Peripyloric area (LN 6, RGEA)	Infrapyloric and suprapyloric lymphadenectomy	Omental incision toward RGEV	Completion of supraduodenal lymphadenectomy
SAU03	Duodenal transection	Division of duodenum	Positioning of duodenum for transection	Hemostasis complete or stapler removed
SAU04	LN5 and RGA	Suprapancreatic dissection, right side	Soft-tissue dissection at suprapancreatic region	Ligation of right gastric artery
SAU05	LN12a	Hepatoduodenal ligament dissection	Completion of RGA ligation	Portal vein exposed at proper hepatic artery
SAU06	LN8	Common hepatic artery dissection	Exposure of common hepatic artery	Soft tissue cleared at CHA
SAU07	LN7, 9 and LGA	Left gastric artery and celiac axis	Surface of LGA revealed	Prior to right crus dissection
SAU08	LN11p	Splenic artery lymphadenectomy	Dissection at splenic artery	Splenic vein exposed, soft tissue cleared
SAU09	LN1 and esophagus	Right crus and paraesophageal dissection	Division of retroperitoneal membrane	Completion of paraesophageal dissection
SAU10	LN3	Lesser curvature dissection	Soft-tissue dissection at lesser curvature	Completion of lesser curvature lymphadenectomy
SAU11	Gastric transection	Division of stomach	Positioning for transection	Hemostasis complete or stapler removed
SAU12	Field maintenance	Pre-reconstruction cleanup	Tidying operative field	Initiation of reconstruction
SAU13	Reconstruction	Billroth-I, Billroth-II, Billroth-II with Braun, or Roux-en-Y	Energy device activation for reconstruction	Anastomosis complete and inspected
SAU14	ICG assessment	Fluorescence imaging	Fluorescence imaging on	Fluorescence imaging off
SFD01	Complete idling	No instrument movement > 20 s	Both operating instruments motionless for > 20 s	Instrument motion resumes or next SAU begins
SFD02	Vision restoration	Lens cleaning or defogging	Camera displacement for cleaning	Completion of camera re-docking

Fourteen semantic action units (SAU) and six surgical flow disruption (SFD) categories, with the start and end criteria used for annotation. The remaining annotated intervals were not defined as SAUs or SFDs and contributed only to total console time and total phase fragmentation.

CHA common hepatic artery, ICG indocyanine green, LGA left gastric artery, LGEA left gastroepiploic artery, LN lymph node, RGA right gastric artery, RGEA right gastroepiploic artery, RGEV right gastroepiploic vein

Code	Phase	Description	Start criterion	End criterion
SFD03	Device cleaning	Instrument cleaning > 20 s	Instrument withdrawn for cleaning, lasting > 20 s	Instrument reintroduced, dissection resumes
SFD04	Re-resection	Additional resection for margin	Initiation of resection beyond the planned transection	Return to the planned workflow
SFD05	Damage repair	Repair of unintended tissue injury	Initiation of repair (suturing, clip application)	Repair complete, workflow returns to the SAU
SFD06	Hemostasis	Control of unexpected bleeding	Hemostasis initiation	Area cleared and hemostasis complete
Annotated intervals not defined as SAUs or SFDs				
—	Preparation	Exploration and preparatory movements (e.g., adhesiolysis)	Preparatory movement before liver traction or dissection	Start of liver traction or greater omentum dissection
—	Liver traction	Retraction of the liver	Insertion of gauze or needle	Activation of instruments for the next phase
—	Specimen retrieval	Placement of the specimen in a bag	Insertion of specimen bag	Closure of bag
—	Final tidying	Field tidying after reconstruction	End of anastomosis check	Camera removed from abdominal cavity
—	Robot docking or undocking	Docking or undocking of the robot	Start of docking or undocking	Completion and move to the next phase
—	Other procedures	Procedures unrelated to the gastrectomy (e.g., cholecystectomy) or not part of the operation itself (e.g., specimen examination, biopsy)	—	—
—	Nonspecific movements	Movements without effect on tissue that break the flow, including instrument changes	—	—

Fourteen semantic action units (SAU) and six surgical flow disruption (SFD) categories, with the start and end criteria used for annotation. The remaining annotated intervals were not defined as SAUs or SFDs and contributed only to total console time and total phase fragmentation.

CHA common hepatic artery, ICG indocyanine green, LGA left gastric artery, LGEA left gastroepiploic artery, LN lymph node, RGA right gastric artery, RGEA right gastroepiploic artery, RGEV right gastroepiploic vein

Block 1 (conventional clinicopathological variables, *n* = 12) included 10 variables: age, sex, BMI, previous abdominal surgery, American Society of Anesthesiologists (ASA) physical status, pathologic T and N categories (American Joint Committee on Cancer, 8th edition), reconstruction type, operative time and estimated blood loss. ASA physical status and pathologic T and N categories were entered as ordinal variables, and reconstruction type as three indicator variables with Billroth-I as the reference, giving 12 candidate predictors. Surgeon identity was not entered as a predictor, because a model containing the three operating surgeons could not be applied to patients operated on by other surgeons.

Block 2 (video-rated surgeon proficiency, *n* = 3) included the three pre-specified sub-scores of the KLASS-02-QC instrument (skill, perigastric and extraperigastric), each calculated by summing the relevant items among the instrument's 22 binary technical items [10, 11]. The item-level sums matched the recorded total score in all 127 operations. Because the sub-scores and the total score are exact sums of the items, only one level of aggregation can be entered into a single model. The

analysis used the three sub-scores, the unit in which the instrument is reported. Entering the 22 binary items would have added many candidates for 22 events, and 9 of the 22 items were achieved in at least 95% of operations (Supplementary Table S1). The sub-scores, unlike the single total score, retain the instrument's three domains.

Block 3 (CT-derived body composition, $n = 4$) was measured on the contrast-enhanced abdominal CT obtained for preoperative staging. On a single axial slice at the level of the third lumbar vertebra (L3), skeletal muscle and adipose tissue were segmented automatically using the Aquarius iNtuition Viewer (version 4.4.13, TeraRecon Inc., Durham, NC, USA) [27, 28]. The Hounsfield unit thresholds were − 29 to + 150 for skeletal muscle and − 190 to − 30 for adipose tissue. The block comprised visceral fat area, subcutaneous fat area and skeletal muscle area (cm^2) and abdominal circumference (cm) at the same level.

Block 4 (video-derived workflow, $n = 42$) was obtained from the console recordings. Each recording was annotated from start to end into phases defined a priori by expert consensus (Table 2). Goal-oriented tasks of the gastrectomy were termed semantic action units (SAUs, 14 phases), and deviations from the expected procedural progression were termed surgical flow disruptions [13, 14] (SFDs, 6 categories). The remaining intervals, such as preparation, liver traction and robot docking, were annotated but not defined as SAUs or SFDs. Timestamps followed consensus-derived criteria and anatomical landmarks, and discrepancies greater than 5 s were resolved by the senior surgeon.

For each SAU and SFD, a duration and a count (number of separate segments) were derived, giving 40 variables. Two global measures, total console time and total phase fragmentation (the length and the number of segments of the whole recording), were added for a total of 42. Because these two measures included the intervals not defined as SAUs or SFDs, they exceeded the summed SAU and SFD values. The four mutually exclusive reconstruction sub-phases were pooled into one phase, and the SAU14 (indocyanine green assessment) count records fluorescence imaging activations. Operational definitions are given in Supplementary Table S2.

In total, the four blocks provided 61 candidate predictors (12, 3, 4 and 42 in blocks 1 to 4, respectively) for 22 complications. The number of events per candidate predictor ranged from 7.33 for block 2 alone to 0.36 for all four blocks combined (Fig. 1). This constraint determined the block-wise design of the study.

Statistical analysis

Patient characteristics were compared across surgeons with the Kruskal–Wallis test and by complication status with the Mann–Whitney U test for continuous variables, and with the chi-square or Fisher's exact test for categorical variables. These descriptive tests were two-sided, and $p < 0.05$ was considered statistically significant.

Durations, counts, operative time and estimated blood loss were $\log(1 + x)$-transformed to limit the influence of a few extreme operations. This transformation estimates no parameters and was applied before the endpoint was examined. Within each resampling split, predictors were standardized using the training fold only, and the same centering and scaling were applied to the held-out fold. There were no missing values in the variables used for analysis.

Each of the 15 non-empty block combinations was fitted using binomial elastic-net regression with the mixing parameter fixed at $\alpha = 0.5$ [29, 30]. The penalty shrinks the coefficients of variables that add little to the fit exactly to zero, which selects variables during fitting. Fixing α kept the penalty identical across combinations, so that differences between combinations reflected the blocks rather than tuning of α. No class-imbalance correction was applied, and the 17.3% event rate was modeled directly.

Discrimination was measured by the area under the receiver operating characteristic curve (AUC) and estimated by repeated nested cross-validation. Five outer folds generated out-of-fold predictions, and five inner folds within each training set selected the penalty parameter (λ.min) by minimizing binomial deviance. The outer cross-validation was repeated 50 times, and we report the median AUC with the 2.5th and 97.5th percentiles. Because these are resampling percentile ranges rather than confidence intervals, they are labeled as 95% ranges throughout. Unless stated otherwise, AUC refers to this nested

estimate. Added predictive value was assessed by comparing the nested AUC of each combination with that of block 1 alone. For the conventional model (block 1 alone), calibration was assessed by comparing predicted and observed risk in quartiles of out-of-fold predicted risk and by the calibration slope and calibration-in-the-large, and the Brier score was used as a summary measure.

For each block combination, the observed nested AUC was compared with an outcome-permutation null distribution, generated by running the identical pipeline on 200 permuted outcomes. Empirical p values were calculated as (1 + the number of null performances greater than or equal to the observed performance) divided by (1 + the number of finite null performances).

To address the secondary aim, each block combination was also fitted once to all 127 patients using the same α and λ-selection rule. Evaluating this full-cohort model in the same patients gave the apparent AUC, and optimism was defined as the apparent minus the nested AUC. Optimism indicates how much performance in the data used for fitting overstates performance in patients not used for fitting. The variables retained by the full-cohort model are reported for every combination, with standardized coefficients for block 1 alone.

The stability of variable selection was assessed for every block combination by stratified half-sampling [31, 32]. In each of 1,000 half-samples, the elastic-net model was refitted using the same α and λ-selection rule, and a variable was considered selected when its coefficient was non-zero. The selection probability (π) was defined as the proportion of half-samples in which a variable was selected. Variables with π above 0.50 were described as stably selected.

Analyses were performed using R 4.3.3 (R Foundation for Statistical Computing, Vienna, Austria) with the glmnet 4.1.10 and pROC 1.19.0.1 packages (random seed 251), and the full analysis code is available on reasonable request (see Code availability). The study is reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) + AI statement [33] where applicable, and the completed checklist is provided in Supplementary Table S3.

Results

Patient characteristics

Of the 127 operations, 22 (17.3%) were followed by a complication of Clavien–Dindo grade II or higher within 30 days (Table 1). Patients with a complication were more often male (86.4% vs. 55.2%, p = 0.007) and more often had ASA physical status III (50.0% vs. 20.0%, p = 0.002), and they had greater estimated blood loss (median 70 vs. 45 mL, p = 0.027). None of the block 2 to 4 measures in Supplementary Table S1 differed significantly by complication status, including the proficiency sub-scores and each item that varied (all p ≥ 0.060), the body composition measures and the two global workflow measures. Complication rates did not differ among the three surgeons (19.5%, 10.0% and 30.0%, p = 0.238, Table 1).

Discrimination across the 15 block combinations

Figure 3a shows the nested AUC of all 15 combinations (Supplementary Table S4), which separated clearly according to whether block 1 was included. The eight combinations including block 1 had nested AUCs of 0.626 to 0.698, whereas the seven without block 1 had nested AUCs of 0.504 to 0.553.

The seven combinations without block 1, which contained only multimodal measures, were not distinguishable from chance, and none differed from its permutation null (p ≥ 0.244). This included proficiency alone (0.512, p = 0.463), body composition alone (0.553, p = 0.244) and workflow alone (0.524, p = 0.463). When these seven models were fitted to all 127 patients, every coefficient was shrunk to zero. None of the 3 proficiency sub-scores, 4 body composition measures or 42 workflow variables was retained. Each model therefore assigned the same predicted risk to all patients, and the apparent AUC was 0.500 (Fig. 3a).

No combination including block 1 achieved a higher nested AUC than block 1 alone. Conventional variables alone yielded a nested AUC of 0.698 (95% range 0.636–0.743, permutation p = 0.015). Adding proficiency, body composition or workflow did not raise this estimate, with nested AUCs of 0.680 (0.606–0.726), 0.684 (0.607–0.725) and 0.647 (0.569–0.715), respectively, and combining all four blocks gave 0.626 (0.510–0.709). Figure 3b shows the corresponding out-of-fold receiver operating characteristic curves for block 1 alone and with each single multimodal block. Seven of the eight combinations including block 1 differed from their permutation null ($p \leq 0.040$), the exception being the combination of all four blocks ($p = 0.055$).

Variables in the conventional model

Block 1 alone, the best-performing combination, retained five of its 12 variables when fitted to all 127 patients. In descending order of absolute standardized coefficient, these were ASA physical status (0.44), male sex (0.28), estimated blood loss (0.15), Billroth-II reconstruction (0.07) and BMI (− 0.05, the only negative coefficient), with π of 0.85, 0.73, 0.58, 0.39 and 0.36, respectively. Among the seven variables not retained, operative time had the highest π (0.23), followed by previous abdominal surgery (0.22), pathologic T category (0.19), pathologic N category (0.19) and age (0.16).

Out-of-fold predicted probabilities were generally consistent with observed risk across the range of predictions (Brier score 0.135, calibration slope 1.18, calibration-in-the-large 0.00). From the lowest to the highest quartile of predicted risk, the predicted probabilities were 8.2%, 12.9%, 18.4% and 29.8%, and the observed complication rates were 3.1% (1/32), 15.6% (5/32), 19.4% (6/31) and 31.3% (10/32).

Optimism and stability of variable selection

Adding the workflow block raised the apparent AUC but not the nested AUC (Fig. 3a, Supplementary Table S4). Among the eight combinations including block 1, the four without the workflow block (12 to 19 candidates) had an apparent AUC of 0.787–0.804 and a nested AUC of 0.664–0.698, corresponding to an optimism of 0.106–0.129. The four with the workflow block (54 to 61 candidates) had an apparent AUC of 0.844–0.874 and a nested AUC of 0.626–0.647, corresponding to an optimism of 0.205–0.227. Block 1 + 4 gave the highest apparent AUC in the study (0.874) but a nested AUC of only 0.647, and its optimism (0.227) was about twice that of block 1 alone (0.106).

Selection stability also decreased as candidates were added, even for conventional variables. Across all 15 combinations, only three variables ever had π above 0.50, namely ASA physical status, sex and estimated blood loss, all of which are conventional variables. The number of such variables fell from three in block 1, block 1 + 2 and block 1 + 3 to two in block 1 + 2 + 3 and block 1 + 4, and to one in each of the three largest combinations. No multimodal variable had π above 0.50 in any combination. The highest values were 0.49 for the number of segments in the greater omentum phase (SAU01) and 0.45 for the duration of the peripyloric area phase (SAU02), both in block 1 + 4. Neither global workflow measure (total console time or total phase fragmentation) was retained in any combination. No body composition measure was retained either, whereas BMI, the simpler measure in block 1, was retained in four combinations, including block 1 + 3.

Discussion

This study showed that, in 127 robotic distal gastrectomies, conventional clinicopathological variables predicted 30-day complications, whereas the added predictive value of multimodal surgical video and body composition data was not yet clearly shown. Combinations without conventional variables were not distinguishable from chance and retained no variable, even when fitted to all patients. Across all combinations, the only variables stably selected were ASA physical status, sex and estimated blood loss, all of which are conventional variables. When performance was estimated in patients not used for model fitting, conventional variables alone gave the highest AUC (nested AUC 0.698), and adding proficiency, body composition or workflow did not raise it (0.647–0.684). In contrast, when the models were evaluated in the same patients used for fitting, every combination that added workflow to conventional variables appeared to outperform conventional

variables alone, and conventional variables plus workflow ranked first. This advantage was an artifact of that evaluation and did not hold in patients not used for fitting.

With advances in analytic methods and the increasing variety of data collected around each operation, the information available for outcome prediction has grown rapidly [7]. Although the operation itself is widely recognized as a determinant of complications, it has remained difficult to convert into variables. To address this, operations have increasingly been divided into phases to quantify the duration and flow of each step [12, 15–18]. Body composition measured on preoperative CT has likewise been associated with complications after gastrectomy [21]. This study therefore explored, in a simple form, whether an established proficiency score, phase-based workflow measures and CT-derived body composition improved prediction when added to conventional clinicopathological variables.

Although all of these multimodal measures were entered, their added value over conventional variables was not clearly shown. One possible explanation is that conventional clinicopathological variables are strong predictors that leave little room for additional measures. Existing models of complications after gastrectomy, including machine-learning models, have relied mainly on routine variables such as age, sex, BMI, ASA physical status and operative time [34–36], the same kind of variables stably selected here. Likewise, no body composition measure was retained in any combination. Visceral fat area tended to be higher in patients with a complication, but part of this difference may have been carried by male sex, which was among the strongest retained variables. Thus, the effects of the multimodal measures may have been smaller than those of conventional variables and too small to detect with 22 events.

It is also possible that the variables used in this study were not optimized to represent the operation or the patient's physique, even if both influence complications. The workflow measures recorded how long each phase took and how often it was interrupted, but not how tissue was handled during that time. The KLASS-02-QC instrument was designed for quality control rather than risk prediction, and its skill sub-score clustered at the maximum of 5 (median 5, interquartile range 5–5). The CT measures were cross-sectional areas not normalized for height, so they did not correspond to established sarcopenia indices. Moreover, the skill and perigastric sub-scores, total console time and total phase fragmentation differed among the surgeons but not by complication status (Supplementary Table S1), suggesting that they reflected who operated more than the patient's risk.

The endpoint may also have limited what video-derived data could show, because it combined general complications such as pneumonia with surgical complications such as anastomotic leakage. General complications are less likely to depend on what happens during the operation, so associations between specific operative steps and specific surgical complications may have been diluted when all complications were analyzed together. The workflow variables closest to stable selection, the number of segments in the greater omentum phase and the duration of the peripyloric area phase, reflect specific operative steps, but their relation to individual complications could not be examined with 22 events.

With so few events, the way performance is estimated can change the conclusion. Adding workflow to conventional variables raised the apparent AUC, estimated in the patients used to build the model, from 0.804 to 0.874, whereas the nested AUC, estimated in patients left out of model building, did not rise. Such overestimation is expected when candidate predictors approach or outnumber events [22, 23], and it occurs whether or not the added variables are truly predictive. A more conclusive test may require a larger number of events [37–40], which multicenter cohorts and automated phase recognition may help to provide [15, 16].

To the best of our knowledge, this study is the first to evaluate whether surgical video, used for both proficiency rating and phase-based workflow, and CT-derived body composition add predictive value to conventional clinicopathological variables after robotic distal gastrectomy. All block combinations were compared under the same conditions, with variable selection confined to the training folds, each estimate compared against a permutation null and selection stability assessed by half-sampling. This design separated true added value from the apparent gain produced by adding candidate predictors. This distinction is easily missed when models are evaluated in the data used to build them.

This study had several limitations. First, the cohort was retrospective and small and came from two centers and three surgeons with unbalanced case numbers (77, 40 and 10 operations), so surgeon and center effects could not be separated. Second, operations were included when a complete, unedited recording was available rather than as a consecutive series, so the cohort may not represent all robotic distal gastrectomies performed during the study period. Third, video rating and phase annotation were performed manually by two reviewers blinded to outcome, with discrepancies adjudicated, but inter-rater agreement was not formally quantified. Fourth, elastic-net regression assumes linear relationships, so nonlinear effects or interactions among variables may have been missed. Finally, no external validation was performed. With 22 events, the study had limited power to detect small improvements, and the 95% ranges, which reflected variation across cross-validation partitions rather than sampling uncertainty, understate the uncertainty around each AUC. Therefore, the absence of a detectable improvement does not prove that none exists.

Despite these limitations, this study was an initial attempt to incorporate surgical video and body composition data into complication prediction after gastrectomy. To clarify the role of these data, further studies in larger cohorts are needed to identify patient-specific and surgical video-specific variables better suited to predicting complications after gastrectomy.

Conclusions

In 127 robotic distal gastrectomies, conventional clinicopathological variables predicted 30-day complications, whereas the added predictive value of multimodal surgical video and body composition data remains to be clearly shown in larger cohorts.

Declarations

Funding

This study was supported by the National R&D Program for Cancer Control through the National Cancer Center funded by the Ministry of Health and Welfare, Republic of Korea (RS-2026-25517509) and by the Korean Gastric Cancer Association (KGCA032023G2). The funders had no role in study design, data collection, analysis, interpretation or the decision to submit.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Conceptualization: J.H.C., H.I.K.; Data curation: H.Y., S.H.P., H.H.L., S.W.L., J.H., M.C., S.H.L., S.H.H.; Formal analysis: H.Y., S.H.P., H.H.L., K.P., W.J., J.H.C.; Funding acquisition: H.I.K.; Investigation: S.H.P., H.H.L., S.W.L., J.H., M.C., S.H.L., S.H.H.; Methodology: H.Y., K.P., W.J., J.H.C., H.I.K.; Project administration: H.I.K.; Resources: W.J.H., J.H.C., H.I.K.; Software: S.H.P., K.P., W.J., J.H.C.; Supervision: Y.M.K., W.J.H., J.H.C., H.I.K.; Validation: J.H., M.C., Y.M.K., W.J.H.; Visualization: S.H.P., H.H.L., J.H.C.; Writing — original draft: S.H.P., H.H.L.; Writing — review and editing: J.H.C., H.I.K. All authors read and approved the final manuscript.

Ethics approval

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. This retrospective study was approved by the institutional review boards of Yonsei University Severance Hospital and Pusan National University Yangsan Hospital (4-2024-0769 and 55-2025-103, respectively). The analysis was not prospectively registered, and no protocol was deposited in a public registry. However, the block definitions, the combination scheme, the elastic-net mixing parameter, the resampling design and the permutation-null procedures were fixed before the outcome was examined.

Consent to participate

The requirement for informed consent was waived by both institutional review boards owing to the retrospective, de-identified nature of the video and clinical data.

Consent to publish

Not applicable. The manuscript contains no individual identifying details, images or video.

Data availability

The de-identified analytic dataset supporting the findings of this study is available from the corresponding author on reasonable request.

Code availability

The analysis code, including data preparation, derivation of the four predictor blocks, modeling of all 15 block combinations, outcome-permutation analyses, variable-selection stability analysis and generation of all figures, is available from the corresponding author on reasonable request. When applied to the analytic dataset with the specified random seed, the code reproduces the reported results.

Patient and public involvement

No patients or members of the public were involved in the design, conduct, reporting or dissemination of this study.

Presentation

This work was presented at the 2025 Annual Meeting of KINGCA, Seoul, Republic of Korea, September 25–27, 2025.

ORCID iDs

Sung Hyun Park 0000-0001-5116-728X; Hannah Heejung Lee 0000-0003-1797-8788; Haemin Yong 0009-0002-4885-7743; Kyungri Park 0009-0007-1737-4809; Woohwan Jung 0000-0003-4561-2214; Sang Won Lee 0000-0001-8348-1062; Jawon Hwang 0009-0008-6759-7426; Minah Cho 0000-0003-3011-5813; Yoo Min Kim 0000-0002-5176-804X; Si-Hak Lee 0000-0002-1923-4777; Sun-Hwi Hwang 0000-0001-8058-0524; Woo Jin Hyung 0000-0002-8593-9214; Jae Hun Chung 0000-0002-7821-5445; Hyoung-Il Kim 0000-0002-6134-4523

References

1. Kim H-H, Hyung WJ, Cho GS, Kim MC, Han S-U, Kim W et al (2010) Morbidity and mortality of laparoscopic gastrectomy versus open gastrectomy for gastric cancer: an interim report–a phase III multicenter, prospective, randomized Trial (KLASS Trial). Ann Surg 251:417–420. https://doi.org/10.1097/SLA.0b013e3181cc8f6b
2. Lee H-J, Hyung WJ, Yang H-K, Han SU, Park Y-K, An JY et al (2019) Short-term Outcomes of a Multicenter Randomized Controlled Trial Comparing Laparoscopic Distal Gastrectomy With D2 Lymphadenectomy to Open Distal Gastrectomy for Locally Advanced Gastric Cancer (KLASS-02-RCT). Ann Surg 270:983–991. https://doi.org/10.1097/SLA.0000000000003217
3. Kim H-I, Han S-U, Yang H-K, Kim Y-W, Lee H-J, Ryu KW et al (2016) Multicenter Prospective Comparative Study of Robotic Versus Laparoscopic Gastrectomy for Gastric Adenocarcinoma. Ann Surg 263:103–109. https://doi.org/10.1097/SLA.0000000000001249
4. Ojima T, Nakamura M, Hayata K, Kitadani J, Katsuda M, Takeuchi A et al (2021) Short-term Outcomes of Robotic Gastrectomy vs Laparoscopic Gastrectomy for Patients With Gastric Cancer: A Randomized Clinical Trial. JAMA Surg 156:954–963. https://doi.org/10.1001/jamasurg.2021.3182

5. Lu J, Xu B-B, Zheng H-L, Li P, Xie J-W, Wang J-B et al (2024) Robotic versus laparoscopic distal gastrectomy for resectable gastric cancer: a randomized phase 2 trial. Nat Commun 15:4668. https://doi.org/10.1038/s41467-024-49013-6
6. Long VD, Thong DQ, Dat TQ, Nguyen DT, Hai NV, Quoc HLM et al (2024) Risk factors of postoperative complications and their effect on survival after laparoscopic gastrectomy for gastric cancer. Ann Gastroenterol Surg 8:580–594. https://doi.org/10.1002/ags3.12780
7. Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A et al (2017) Surgical data science for next-generation interventions. Nat Biomed Eng 1:691–696. https://doi.org/10.1038/s41551-017-0132-7
8. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR et al (2013) Surgical skill and complication rates after bariatric surgery. N Engl J Med 369:1434–1442. https://doi.org/10.1056/NEJMsa1300625
9. Woods MS, Liberman JN, Rui P, Wiggins E, White J, Ramshaw B et al (2023) Association between Surgical Technical Skills and Clinical Outcomes: A Systematic Literature Review and Meta-Analysis. JSLS 27. https://doi.org/10.4293/JSLS.2022.00076. e2022.00076
10. Han S-U, Hur H, Lee H-J, Cho GS, Kim M-C, Park YK et al (2021) Surgeon Quality Control and Standardization of D2 Lymphadenectomy for Gastric Cancer: A Prospective Multicenter Observational Study (KLASS-02-QC). Ann Surg 273:315–324. https://doi.org/10.1097/SLA.0000000000003883
11. Song JH, Shin HJ, Hyung WJ, Yang H-K, Han S-U, Park Y-K et al (2023) Predictive Value of KLASS-02-QC Assessment Score on KLASS-02 Surgical Outcomes: Validation of Surgeon Quality Control and Standardization for D2 Lymphadenectomy. Ann Surg 278:e1011–e1017. https://doi.org/10.1097/SLA.0000000000005810
12. Garrow CR, Kowalewski K-F, Li L, Wagner M, Schmidt MW, Engelhardt S et al (2021) Machine Learning for Surgical Phase Recognition: A Systematic Review. Ann Surg 273:684–693. https://doi.org/10.1097/SLA.0000000000004425
13. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt TM (2007) Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. Surgery 142:658–665. https://doi.org/10.1016/j.surg.2007.07.034
14. Parker SEH, Laviana AA, Wadhera RK, Wiegmann DA, Sundt TM (2010) Development and evaluation of an observational tool for assessing surgical flow disruptions and their impact on surgical performance. World J Surg 34:353–361. https://doi.org/10.1007/s00268-009-0312-z
15. Takeuchi M, Kawakubo H, Tsuji T, Maeda Y, Matsuda S, Fukuda K et al (2023) Evaluation of surgical complexity by automated surgical process recognition in robotic distal gastrectomy using artificial intelligence. Surg Endosc 37:4517–4524. https://doi.org/10.1007/s00464-023-09924-9
16. Komatsu M, Kitaguchi D, Yura M, Takeshita N, Yoshida M, Yamaguchi M et al (2024) Automatic surgical phase recognition-based skill assessment in laparoscopic distal gastrectomy using multicenter videos. Gastric Cancer 27:187–196. https://doi.org/10.1007/s10120-023-01450-w
17. Komatsu M, Kitaguchi D, Ninomiya Y, Takabe Y, Yoshida M, Yura M et al (2026) Artificial Intelligence–Based Surgical Phase Analysis Enables Objective Assessment of Surgeon Skill in Robotic Distal Gastrectomy: A Multicenter Study. https://doi.org/10.1002/ags3.70249. Ann Gastroenterol Surg
18. Umemiya A, Takenaka S, Kitaguchi D, Arakaki S, Wakabayashi M, Iuchi S et al (2026) AI-based assessment of surgical efficiency and exposure quality in laparoscopic distal gastrectomy using a phase recognition model. Surg Endosc 40:8751–8762. https://doi.org/10.1007/s00464-026-13188-4
19. Tweed TTT, van der Veen A, Tummers S, van Dijk DPJ, Luyer MDP, Ruurda JP et al (2022) Body Composition Is a Predictor for Postoperative Complications After Gastrectomy for Gastric Cancer: a Prospective Side Study of the LOGICA Trial. J Gastrointest Surg 26:1373–1387. https://doi.org/10.1007/s11605-022-05321-0
20. Imai Y, Lee S-W, Kawai M, Tashiro K, Kawashima S, Tanaka R et al (2022) Visceral fat area is a better indicator of surgical outcomes after laparoscopic gastrectomy for cancer than the body mass index: a propensity score-matched analysis. Surg Endosc 36:3285–3297. https://doi.org/10.1007/s00464-021-08642-4

21. Ren T, Peng Y, Ma L, Chen Q, Jing Y, Xu Y et al (2026) Association of computed tomography-derived body composition and complications after radical gastrectomy for gastric cancer: A systematic review and meta-analysis. Eur J Radiol 204:113154. https://doi.org/10.1016/j.ejrad.2026.113154
22. Harrell FE, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15(19960229):361–387. https://doi.org/10.1002/(SICI)1097-0258.)15:4%3C361::AID-SIM168%3E3.0.CO;2-4
23. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 54:774–781. https://doi.org/10.1016/s0895-4356(01)00341-9
24. Ambroise C, McLachlan GJ (2002) Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 99:6562–6566. https://doi.org/10.1073/pnas.102102699
25. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7:91. https://doi.org/10.1186/1471-2105-7-91
26. Dindo D, Demartines N, Clavien P-A (2004) Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg 240:205–213. https://doi.org/10.1097/01.sla.0000133083.54934.ae
27. Cho S, Shin S, Lee S, Rhee Y, Kim H-I, Hong N (2024) Differential Impact of Subcutaneous and Visceral Fat on Bone Changes after Gastrectomy. Endocrinol Metab (Seoul) 39:632–640. https://doi.org/10.3803/EnM.2024.1956
28. Jeong JY, Hwang J, Park SH, Cho M, Kim YM, Hyung WJ et al (2026) Simple clinical parameters to identify sarcopenia 1 year after gastrectomy for gastric cancer. Gastric Cancer 29:841–853. https://doi.org/10.1007/s10120-026-01748-5
29. Zou H, Hastie T (2005) Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B Stat Methodol 67:301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
30. Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33:1–22. https://doi.org/10.18637/jss.v033.i01
31. Meinshausen N, Bühlmann P (2010) Stability Selection. J R Stat Soc Ser B Stat Methodol 72:417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x
32. Shah RD, Samworth RJ (2013) Variable Selection with Error Control: Another Look at Stability Selection. J R Stat Soc Ser B Stat Methodol 75:55–80. https://doi.org/10.1111/j.1467-9868.2011.01034.x
33. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B et al (2024) TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385:e078378. https://doi.org/10.1136/bmj-2023-078378
34. Kunisaki C, Miyata H, Konno H, Saze Z, Hirahara N, Kikuchi H et al (2017) Modeling preoperative risk factors for potentially lethal morbidities using a nationwide Japanese web-based database of patients undergoing distal gastrectomy for gastric cancer. Gastric Cancer 20:496–507. https://doi.org/10.1007/s10120-016-0634-0
35. Lin Z, Yan M, Chen H, Wei S, Li Y, Jian J (2025) Development and validation of a machine learning model to predict postoperative complications following radical gastrectomy for gastric cancer. Front Oncol 15:1606938. https://doi.org/10.3389/fonc.2025.1606938
36. Lee S, Oh HJ, Yoo H, Kim CY (2025) Machine learning insight: unveiling overlooked risk factors for postoperative complications in gastric cancer. Cancers (Basel) 17:1225. https://doi.org/10.3390/cancers17071225
37. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG et al (2019) Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 38:1276–1296. https://doi.org/10.1002/sim.7992
38. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 49:1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3

39. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB et al (2020) Calculating the sample size required for developing a clinical prediction model. BMJ 368:m441. https://doi.org/10.1136/bmj.m441
40. van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ et al (2019) Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res 28:2455–2474. https://doi.org/10.1177/0962280218784726

Figures

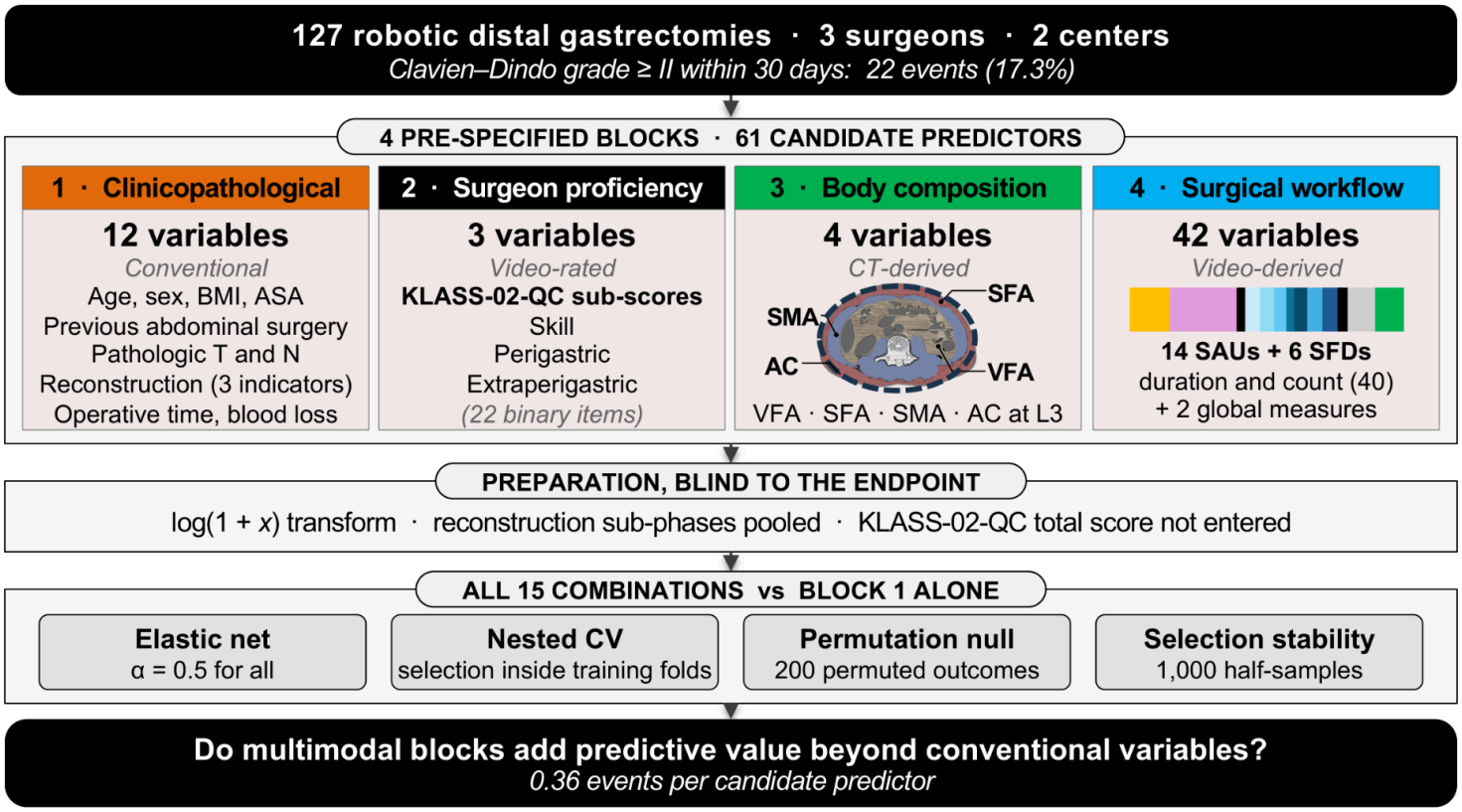

Figure 1

Study design. AC abdominal circumference, ASA American Society of Anesthesiologists physical status, BMI body mass index, CT computed tomography, CV cross-validation, KLASS-02-QC Korean Laparoendoscopic Gastrointestinal Surgery Study Group 02 quality control, L3 third lumbar vertebra, SAU semantic action unit, SFA subcutaneous fat area, SFD surgical flow disruption, SMA skeletal muscle area, VFA visceral fat area

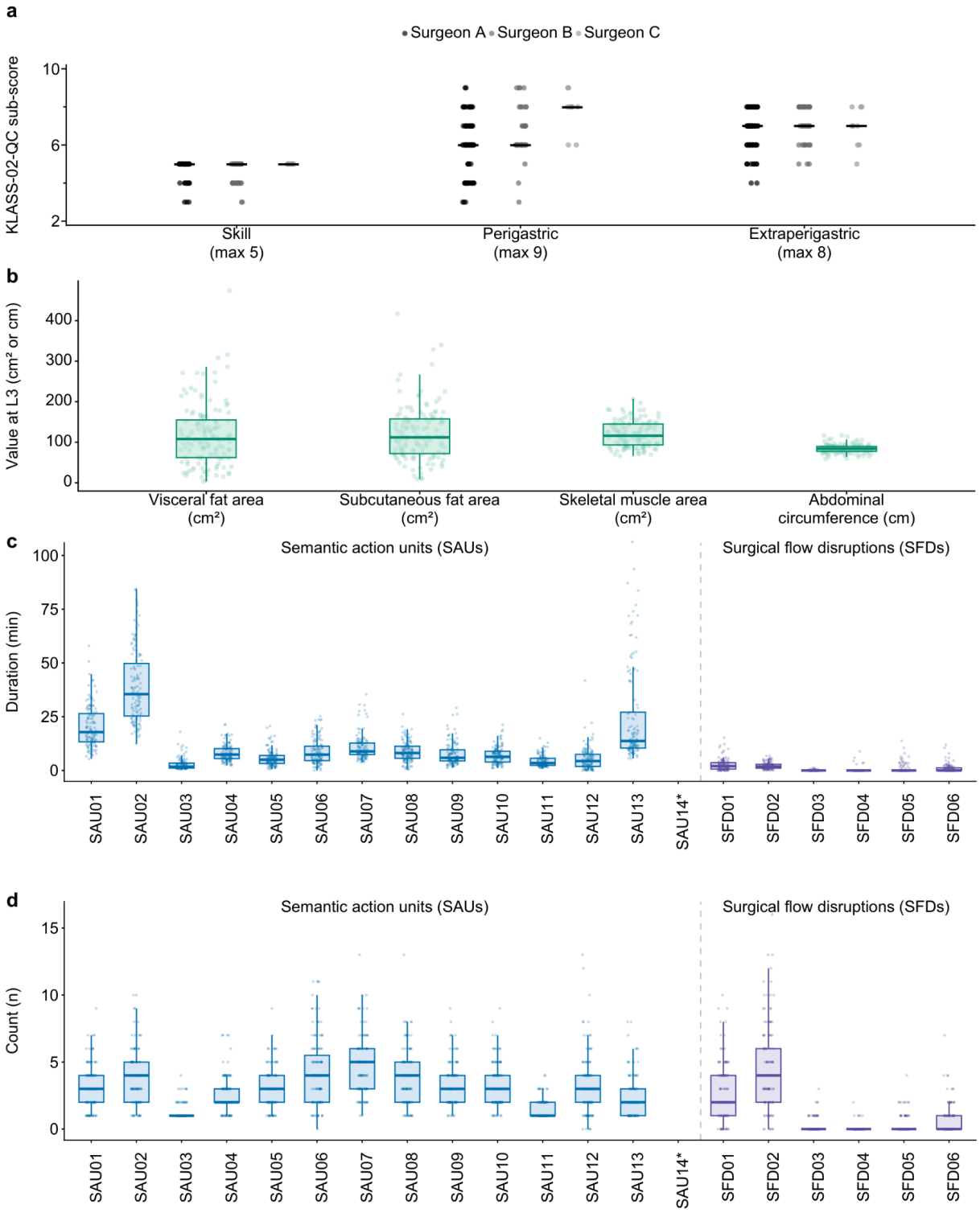

Figure 2

Distribution of blocks 2 to 4. **a** Proficiency sub-scores by surgeon (bars, medians). **b** Body composition at L3. **c, d** Duration and count of each SAU and SFD (codes in Table 2). Boxes, medians and interquartile ranges. Whiskers, 1.5 times the interquartile range. *SAU14 is omitted from c and d because its count records imaging activations

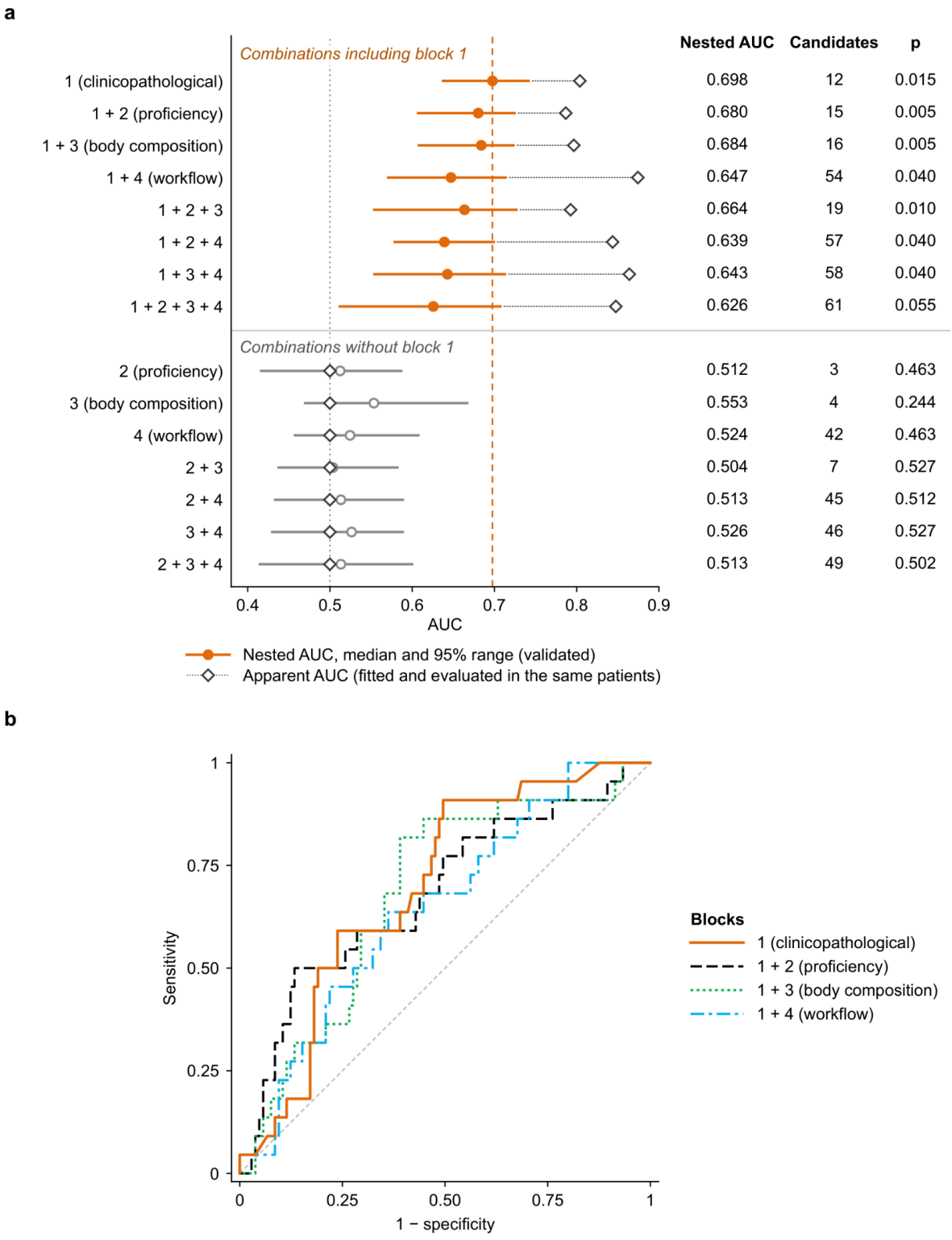

Figure 3

Discrimination for 30-day complications. **a** Nested AUC for the 15 block combinations. Filled orange circles, medians for combinations with block 1. Open grey circles, medians for combinations without block 1. Bars, 2.5th–97.5th percentiles across 50 repeats. Diamonds, apparent AUC. Dashed line, block 1 alone. Dotted line, AUC of 0.5. Candidates, number of predictors entered. p, permutation test. **b**Receiver operating characteristic curves from the repeat closest to the median AUC. AUC area under the curve

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf