

Supplemental table 1: Complete search algorithms

Database	Search algorithm
PubMed	(surgeon* OR "Surgeons"[Mesh]) AND (coaching OR coach* OR "Mentoring"[Mesh] OR "Work-Life Balance"[mesh] OR "Work Performance"[mesh] OR "self care"[Mesh] OR "Formative Feedback"[mesh] OR "Burnout, Professional"[mesh]) AND (English [lang]) NOT (Comment [PT] OR Editorial [PT] OR Letter [PT] OR News [PT] OR Review [PT])
Scopus	TITLE-ABS KEY ((surgeon*) AND (coach* OR mentoring OR "Work-Life Balance" OR "Work Performance" OR "self care" OR "Formative Feedback" OR burnout)) AND DOCTYPE (ar) AND (LIMIT-TO (LANGUAGE , "English"))
Web of Science	(((surgeon*) AND (coach* OR mentoring OR "Work-Life Balance" OR "Work Performance" OR "self care" OR "Formative Feedback" OR burnout))) DOCUMENT TYPES: (ARTICLE) AND LANGUAGES: (ENGLISH)
Clinicaltrials.gov	(Surgeon OR surgery) AND coaching
Cochrane Central Register of Controlled Trials	(Surgeon OR surgery) AND coaching
WHO International Clinical Trials Registry Platform Search Portal	(Surgeon OR surgery) AND coaching

Supplemental table 2: Data Extraction Form

Article number	
Title	
Authors	
Author affiliations	
Date	
Journal	
Population	
Eligibility criteria for coachees? (sources/methods of selection)	
Recruitment period (dates)	
Surgeons' specialty?	
Level of experience of surgeon (how many years of practice, age, level of expertise, etc)	
Number of coachees (surgeons)?	
Were coachees from same or different services, different hospitals?	
Setting of intervention (city and country)?	
Intervention	
Study Goal	
Definition of Surgical Coaching	
Who were the coaches? (surgeons, peers, professionals)	
Coaching Intervention Goal	
Number of coaches	
Who were the coaches?	
Were the coaches trained?	
Interaction type (In-person, web based, video conference?)	
Setting of coaching sessions	
Number of coaching sessions	
Duration of coaching sessions	
Description of intervention and coaching	
Assessment, metrics, or indicators used?	
Improvement solutions used?	
Outcomes	
Surgeons' reactions about being coached	
Surgeon non-technical skills	
Surgeon's well-being	
Patient adverse outcomes	
Study Design	
Case-control, pre-post, Randomized Control Trial, Factorial design, cross over, etc.?	

GRADE (very low, low, moderate, high)	
Main findings	
Reported findings	
Results (statistics used, time points reported, etc.)	
Limitations	
Was the program effective	
Implementation	
Reported barriers (related to the coach, coachee, coaching program, context, etc.)	

Supplemental table 3: Coachee population

	Recruitment methods, sources, and eligibility criteria for coachees	Surgeons' specialty	Level of experience of surgeons	Number of surgeons	Were coachees from same or different departments, different hospitals?
Greenberg et al. 2018	Coached surgeons were solicited from the Wisconsin Surgical Society (state chapter of the American College of Surgeons)	Multiple specialties (6 general surgeons; 6 sub specialties)	Participating surgeons varied in experience (6 with <5, 4 with 5–15, and 2 with >15-years of experience)	12 surgeons completed the introductory session, but only 11 completed the video-based coaching sessions (no controls)	Different hospitals across Wisconsin
Bull et al. 2020	Residents, fellows, and attending surgeons from a single institution who performed laparoscopic general surgery were recruited. Participants were required to perform laparoscopic surgery as the primary operator for inclusion. First year residents were excluded.	Surgeons who performed laparoscopic general surgery (Appendectomy, cholecystectomy, inguinal hernia repair, and colonic resection were the most common laparoscopic cases chosen for rating)	Fellows and attending surgeons (1-18 years after residency) as well as residents performing laparoscopic general surgery as the primary operator were invited to participate in the study. (First year residents were excluded)	12 total: 5 training residents, 3 fellows, and 4 attending surgeons (no controls)	Single center
Maynard et al. 2020	Surgeons were sampled from a large teaching hospital in the Rocky Mountain Region of the United States. 20 surgeons were chosen to receive coaching. To be included, surgeons had to have at least one surgery observed and coded before and after the coaching intervention.	N/A	N/A	20 surgeons received coaching (20 quasi-control surgeons did not receive coaching)	Surgeons were from different departments of one large teaching hospital in the Rocky Mountain Region.
Pradarelli et al. 2020	Recruitment was conducted via email invitation based on the champions' professional networks (The champion was the representative from each study center). No instructions were provided to the champions regarding recruitment or assignment of coaching roles based on surgeons' age, sex, clinical experience, specialty, or academic position.	Orthopedic, otolaryngology, urology, and general surgery subspecialties	Most coachees were in their first decade in practice (61%), although three coachees had over 20 years of clinical experience.	23 coachees (no controls)	4 participating medical centers – a faculty surgeon at each was in charge of recruiting surgeon participants
Sitzman et al. 2020	Attending plastic surgeons in the United States or Canada from the Ameri-cleft Task Force Surgeon Subgroup who performed 10 or more cleft palate repairs annually were contacted by the primary investigator for recruitment.	Plastic Surgery	Attending surgeons (no additional information about experience level given)	2 surgeons received coaching (3 were assigned to control group)	6 different institutions
Greenberg et al. 2021	Any surgeon member of the Michigan Bariatric Surgery Collaborative (MBSC) with 2 years of data on a minimum of 10 procedures before and after the coaching intervention was eligible.	Bariatric Surgery	Surgeons had an average of 12.5 years of experience in bariatrics.	26 surgeon coachees, of which 6 also served as peer coaches. (25 “non-participants” did not receive coaching)	Different institutions across Michigan (The Video Analysis and Peer Coaching in Bariatric Surgery program was a statewide coaching initiative of the MBSC)

Duclos et al. 2020	Professional surgeons from the Hospices Civils de Lyon were recruited. To be eligible, surgeons must be present during the three years of the study, be below the age of retirement (<64), and complete more than 50 surgeries per year.	Digestive, orthopedic, gynecology, urology, cardiac, and thoracic surgery	Faculty/attending surgeons (no surgeons in training were included)	20 coachees (20 control)	13 different surgery departments across 5 hospitals in Lyon, France
Dyrbye 2020	Surgeons at the Mayo clinic were recruited to participate in the study.	N/A	N/A	80 total coachees (crossover design)	Mayo clinic

Supplemental table 4: Intervention

Study ID	Coaches	Details about the coaches	Details of coach training	Setting of coaching sessions	Number of coaching sessions	Duration of coaching sessions	Detailed description of intervention and coaching	Reproducible improvement solutions used?	Metrics/indicators used? If yes, what metrics and/or indicators used?
Greenberg et al. 2018	Surgeons (peer)	Peer nominations were solicited from the membership of Wisconsin surgical society to identify surgeons with the technical and interpersonal skills to serve as coaches. All surgical coaches had at least 15 years of experience in their specialty (3 general surgeons, 5 subspecialists) and worked in a range of practice settings (2 academic, 5 urban private practice, 1 rural private practice).	4-hour coach training session, coach training manual, tools, and instructional videos, were developed to guide the coaches through the process.	Face to face, within 1 month after the filmed operation.	1 to 3	Average length = 56 minutes (range = 21-96)	Coaching pairs were matched based on participant preferences, specialty, practice setting, area of focus for coaching, and geography. Surgeons were not paired with other surgeons from their own institution The pairs began with an introductory session that was designed to establish rapport, set goals, and develop an action plan for the coaching sessions to follow. Surgeons then audio- video recorded operations of their choice for a series of 1-hour video review sessions with their coach. The coach-participant pairs then met for a coaching session within 1 month of the recording. During coaching sessions, pairs were expected to review and refine the participant's goals, analyze the participant's video-recorded surgical procedure, identify changes to implement in practice, and develop a plan for enacting those changes before the next coaching session	No	No
Bull et al. 2020	Surgeons (trained, investigators of study)	The same 2 study investigators performed the coaching. The senior coach was an experienced general surgeon and researcher in the domains of surgical education, performance assessment, and coaching. The junior coach was a surgical resident who was formally instructed in rater training and coaching techniques prior to commencement of the study.	Trained in coaching techniques and as 'expert raters' using OSATS. No strict coaching models were applied. Coaches applied specific techniques based on the discussion content, the existing coaching relationship, feedback from participants, and personal experience. The GROW or modified PRACTICE models were most commonly used.	Face to face (in private), 3-14 days post-surgery.	3 to 5	about 90 minutes	An individualized coaching intervention was performed for each participant. Participants were instructed to identify up to 5 videos for self-assessment. The coaching intervention was performed over 3 to 5 sessions. 1-2 personal videos were discussed during each coaching session. The first session included an introduction to coaching. The following sessions were loosely structured around case debriefing and the review of one or more participant videos. The two coaches independently identified key moments during the footage or general themes for discussion during the upcoming session. They then compared coaching notes before meeting with the participant. Each session was performed 3-14 days after the operation. Further material, such as example videos or references in the literature, were used to supplement discussion and prepared beforehand.	No	Yes, each participant reviewed the same peer video at baseline, at the time of the third self-assessment, and at completion of the intervention. Peer-assessment videos were included to identify variation in rating accuracy over time. They provided a point of reference for improving use of the OSATS scale.
Maynard et al. 2020	Surgeon (trained, retired)	A retired orthopedic surgeon certified in crew resource management training (CRM)	The coach was a retired surgeon and trained specialist trained in CRM training	Face to face. Immediately after operation.	Team leader coaching session immediately after each surgery conducted during the study period (57 surgeries total, an average of 2.85 per surgeon)	N/A	The coach observed all surgeons performing surgeries during the baseline period, noting how well each promoted teamwork and potential improvement areas. He later observed surgeries completed by the 20 surgeons in the coaching intervention condition. Immediately after each surgery, he conducted team leader coaching sessions, providing the surgeon with feedback regarding what they had done well and what could have been improved upon to better facilitate effective teamwork. These sessions included discussions of CRM principles and examples of what he had observed during the surgery to reinforce CRM principles. Coaching sessions were nonstandardized, and guided by the needs, skills, and experiences of the individual being coached.	No	Yes, Three trained subject-matter expert observers (SMEs; i.e., a retired surgical nurse, former surgical equipment representative, and healthcare administration doctoral-level student) assessed team processes and outcomes before and during surgeries at baseline and Five months after the coaching intervention.

Pradarelli et al. 2020	Surgeons (peer)	Professional surgeons (The champions from each medical center assigned their colleagues as either "coachee" or "coach" for the duration of the program and paired coachees 1:1 with a coach. The coach/coachee matching process was not standardized for the inaugural SCOPE program).	Coaches attended a 3-h, in-person training to learn the core principles and skills of surgical coaching. This was followed by weekly emails to remind surgeons of coaching tips from the training	Face to face (in the days prior to operation (for preoperative goal setting) and Within 0 to 72 hours after real life observation of operation.	3	N/A	Preoperative goal-setting occurred within days preceding the operation. Intraoperatively, coaches were expected to directly observe their coachee without scrubbing in, acting as a "fly on the wall" and observing activity related to the coachee's goals. Postoperative debriefs were completed within 3 days following the case. Coach/coachee pairs could discuss any aspect of intraoperative performance during their coaching sessions. For guidance, participants were provided with language to discuss technical skills (i.e., respect for tissue, exposure, instrument handling, time and motion, flow of operation), non-technical skills (i.e., situation awareness, decision making, communication and teamwork, leadership), or intraoperative teaching skills (i.e., teaching the operation, teaching new surgical knowledge, assessing comprehension).	No	Yes, OSATS for technical skills, NOTSS for non-technical
Sitzman et al. 2020	Surgeons (peer)	Other surgeon participants in the study with a fistula incidence at or below the study's mean.	No	Face to face, over the course of the 2-day visit.	2-day visit with coaching	N/A	All participants received feedback based on 3-68 months of baseline audit. Surgeons were then either assigned to the coaching group, or a group that was invited to make self-directed changes. The coachees went to their assigned coach's hospital for a 2-day visit during which they reviewed video of the coachee conducting surgery during the baseline period. The coachee directly observed the coach during at least one operation, during clinic, and during ward rounds. The coach helped the coachee develop a personalized action plan containing up to three specific changes in their surgical technique or perioperative care.	No	Yes, Audit and feedback – collects standardized outcome measurements and reports individuals and peer group results.
Greenberg et al. 2021	Surgeons (peer)	15 surgeons from the MBSC who had the lowest rates of serious complications in the two years (2013-2014) prior to the study were invited to serve as surgical coaches. 14 of the 15 surgeons accepted and were trained in the principles of surgical coaching.	4-hour session facilitated by a board-certified surgeon and a certified executive coach. Following formal training, each coaching session included a "refresher" in coaching principles.	one-on-one. The sessions were co-located and immediately preceded the MBSC meetings every 4 months	Median of 4 sessions (range = 1-8). 9 sessions were offered, surgeons could participate in as many as desired	1 hour	The participants provided videos of their own sleeve gastrectomy procedures. Videos were viewed on a laptop computer for the coaching interactions, which took place immediately before the MBSC meetings held every four months.	No	No
Duclos et al. 2020	Professional	Professional(s)	Professionally trained	Face to face	N/A	N/A	N/A	Yes, Modules of improvement will be proposed	Yes, Charting system providing feedback on patient outcomes and profiling of individual surgeons
Dyrbye 2020	Professional	Professional	Professionally trained	One-on-one over phone	6	N/A	N/A	N/A	N/A

Reproducible improvement solutions – strategies, solutions, or tools that can be used for all participants and reproduced in future studies

Supplemental table 5: Findings

Study ID	Reported findings	Quantitative results	Qualitative evidence reported ?	Study limitations	Reported barriers to implemented future coaching programs
Greenberg et al. 2018	This study was the first report on cross-institutional surgical coaching for the continuous professional development of practicing surgeons. Findings showed perceived value among participants. Logistical challenges for implementing this evidence-based program were identified.	Mean objective ratings of coach effectiveness was 3.1 ± 0.7 , ranging from 2.0 to 5.0 on specific activities of coaching. Subjective ratings by coaches and participants were consistently higher. Coaches reported that the training provided effectively prepared them to facilitate coaching sessions. Participants were similarly positive about interactions with their coaches. Likert scale ratings showed that participants were satisfied with their experience (mean 4.4 ± 0.7) and found the coaching program valuable (mean 4.7 ± 0.7).	Yes	Lack of control group	Audio-video technology for video capture in the OR and scheduling of sessions (Time). Inconsistency in defining and conception of coaching was identified as a barrier for future implementation of surgical coaching
Bull et al. 2020	Targeted coaching using video review of laparoscopic cases improved operative self-assessment accuracy using the OSATS scale. The post-intervention questionnaire showed that there was strong support for video review and coaching using structured feedback. All participants agreed that this process should be a regular component of surgical practice.	At baseline, there was no correlation between self-assessment and expert ratings. After completion of the coaching program there was correlation between self-assessment and expert ratings ($P = .003$) and improved self-assessment accuracy compared to baseline ($P = .041$).	No	The study's single location limits the generalizability of the results. The study also included residents, fellows, and faculty so we cannot isolate the effect on faculty.	Time constraints and receptiveness to feedback were reported as potential barriers to surgical coaching.
Maynard et al. 2020	Results from a multilevel mixed-model (treatment vs. control, over time) structural equation model suggest that teams where the surgeon (team leader) received the coaching intervention exhibited higher-quality team transition processes. Transition processes related positively to subsequent action and interpersonal processes, which in turn yielded improvements in two different surgical team performance outcomes.	After controlling for co-variates, the coaching intervention related positively to transition processes, which related positively to action and interpersonal processes.	No	Unknown whether these findings can be generalized to other contexts.	None reported
Pradarelli et al. 2020	Changes in coaches' technical and non-technical skills were not detected over three coaching sessions. Considered longitudinal peer surgical coaching to be a meaningful strategy for surgeons' professional development.	Adjusted mean OSATS ratings did not vary over three coaching sessions (4.39 vs 4.52 vs 4.44, respectively; $P = 0.655$). Adjusted mean total NOTSS ratings also did not vary over three coaching sessions (15.05 vs 15.50 vs 15.08, respectively; $P = 0.529$).	Yes	Fewer than half of the subjects completed all three coaching sessions	Surgeons' limited time was reported as a barrier.
Sitzman et al. 2020	Surgeon-delivered audit and feedback incorporating peer coaching on technical performance was feasible for surgeons	Seven surgeons enrolled in the trial. All seven completed the baseline audit and disclosed their fistula incidence to other participants. The median baseline fistula incidence was 0.4 percent (range, 0 to 10.5 percent). Two surgeons were unable to receive the feedback intervention. Of the five remaining surgeons, two were allocated to intensive feedback and three to simple feedback. All surgeons completed their assigned	Yes	Very small sample size and low operative volume coupled with low fistula incidence.	Reported barriers included difficulty scheduling in-person coaching sessions when participants are primed to receive feedback.

		feedback intervention. Among surgeons receiving intensive feedback, fistula incidence was 5.9 percent at baseline and 0.0 percent following feedback (adjusted OR, 0.98; 95 percent CI, 0.44 to 2.17).			
Greenberg et al. 2021	Surgeon coachees' operative times improved, but there was no significant improvement in risk-adjusted outcomes.	The program was not associated with significant change in risk-adjusted complications with relative risks for coaches, participants and non-participants of 0.99 (0.62-1.37), 0.91 (0.64-1.17) and 1.15 (0.83-1.47), respectively. Operative times did improve for participants, but not coaches or non-participants, with risk differences of -14.0 (-22.3, -5.7), -1.0 (-4.5, 2.4), and -2.6 (-6.9, 1.7).	Yes	Study findings may not be generalizable to other surgery specialties.	Potential perception of hierarchical relationship between coach and coachee during peer coaching.
Duclos et al. 2020	N/A	N/A	N/A	N/A	N/A
Dyrbye 2020	N/A	N/A	N/A	N/A	N/A

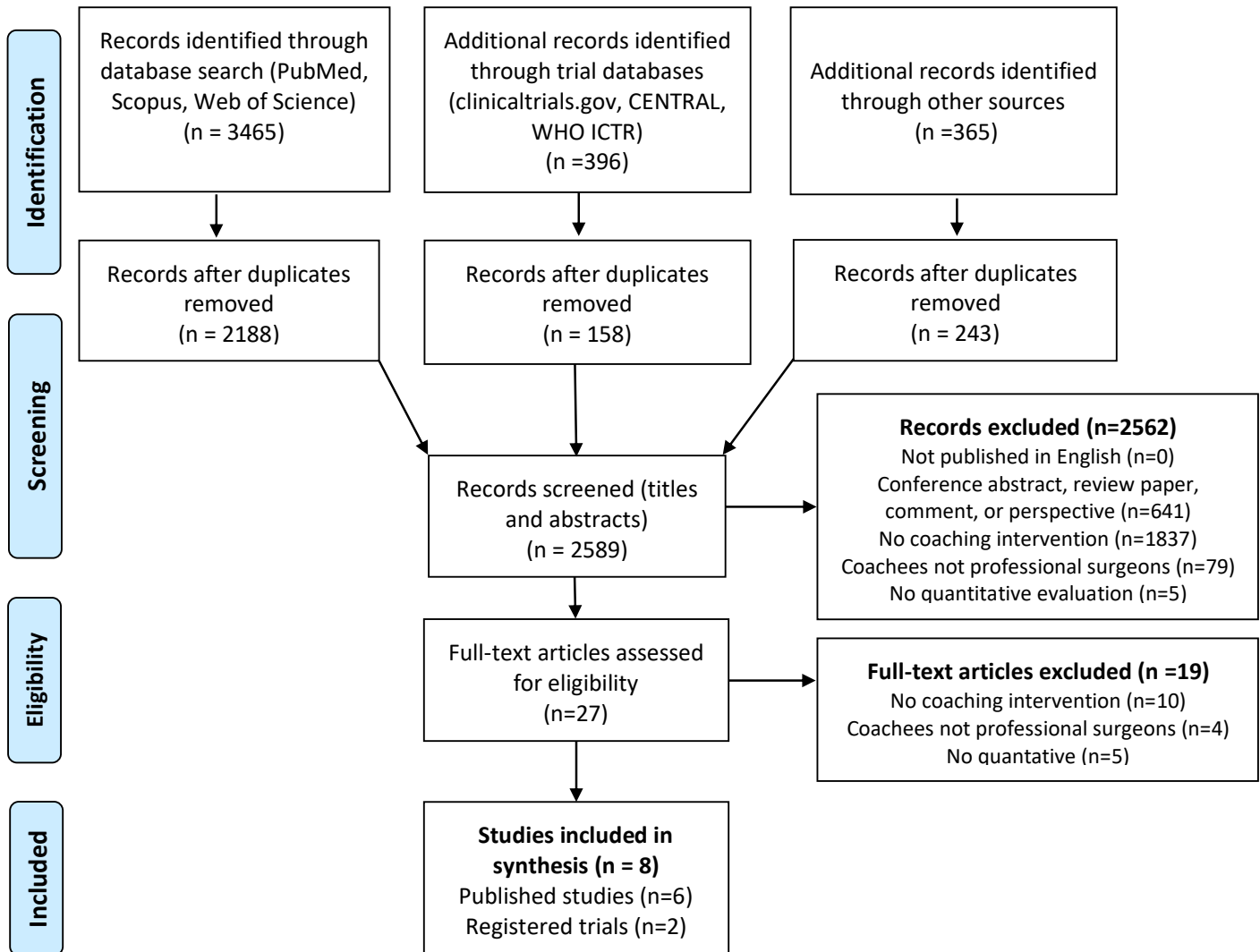
OSATS - Objective structured assessment of technical skills scale

NOTTS - Non-Technical Skills for Surgeons rating scale

Supplemental Figure 1



PRISMA Surgical Coaching Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.