| Paper | Final Paper Number | Authors | Notes | Cohort Size | Did the trial have a clearly focussed issue? | Was assignment of patients to treatments randomized? |
|-------|--------------------|---------|-------|-------------|--------------------------------------------|--------------------------------------------------|
| 4     | 1                  | Woofter M, et al. | Carotid endovascular stenting, patients anatomy uploaded to simulator and practised on before surgery | 15 (n=9, i=6) | Yes: does rehearse using patient specific anatomy uploaded to a simulator improve procedural efficiency and outcomes? | Yes |
| 8     | 2                  | Meertens H, et al. | Endovascular interventions in lower extremity | 32, 3 dropouts (n=10, i=10, 2=9) | Yes: how does the PROSPECT training program compare to e-learning and traditional training with respect to acquisition of endovascular skills and the transferability of these skills to real patient scenarios? | Can't tell |
| 11    | 3                  | Zavlin B, et al. | Bariatric surgery, practice and assessment on pigs, intervention group rated on live patient after trial | 20 (n=10, i=10) additionally 9 chief residents | Yes: to develop and provide evidence of validity for a simulation-enhanced training curriculum for an advanced minimally invasive surgical procedure | Can't tell |
| 14    | 4                  | Alexandcr L, et al. | Anurans repair: half of patients anatomy uploaded to simulator and practised on before actual operation | 100 (n=50, i=50) | Yes: to evaluate the effect of patient specific rehearse prior to endovascular aneurysm repair on patient safety and procedural efficacy | Can't tell |
| 19    | 5                  | Nikkola C, et al. | Simulation of endoscopic camera navigation. Simulation text and transfer test in OR, navigating camera in a cholecystectomy | 36 (n=12, i=11 (1 dropout), k = 12) | Yes: how to train laparoscopic camera navigation and the transfer of this training to the operating room | Can't tell |
| 24    | 6                  | Waterman BR, et al. | Diagnostic shoulder arthroscopy | 22 (n=10, i=12) | Yes: dose simulation training on diagnostic shoulder arthroscopy improve task performance in the operating room | Can't tell |
| 25    | 7                  | Shore EM, et al. | Salpingectomy and intracorporeal knot tying, assessment in OR | 27 (n=13 (2 excluded from primary outcome measures), i=14 (4 excluded from primary outcome measures)) | Yes: to develop and validate a comprehensive ex-vivo training curriculum for gynaecological laparoscopy | Can't tell |
| 36    | 8                  | Patel NR, et al. | Salpingectomy, porcine simulation model, human pre and post test | 22 (n=11, i=11) | Yes: To evaluate the effectiveness of the porcine training model for OB/GYN residents in laparoscopic salpingectomy | Can't tell |
| 43    | 9                  | Dunn JC, et al. | Arthroscopy of shoulder joint. In vivo test, then simulation, then another in vivo test. I year later new in vivo test to look at skill decay | 17 (they legitimately dont report how many were in each group) | Yes: are gains made by residents after a simulation program retained after a period away from training? | Can't tell |
| 46    | 10                 | Pullen ID, et al. | Central venous catheter placement | 73 (n=37, i=36), observed CVC placements=87 (n=38, i=49) | Yes: to evaluate the effects of a programatic CVC simulation program on procedural protocol adherence, technical skill and patient outcomes | Can't tell |
| 49    | 11                 | Glover SG, et al. | Colonoscopy, simulation pre-test, training, post-test then in-vivo post-test | 34, 1 dropout (n=17, i=16) control = self-regulated learning, intervention = structured curriculum | Yes: to determine whether structured simulation based curriculum improves performance in colonoscopy and transfers to the clinical setting as compared to a self regulated curriculum | Can't tell |
| 53    | 12                 | Carlsten CG, et al. | Inguinal hernia repair: Mincequin and pig simulation then in-vivo testing | 18 (n=9, i=10), included in final analysis (n=7, i=9) | Yes: to test the effect of a module-based training model in Lichtenstein hernia repair technical skills | Can't tell |
| 56    | 13                 | Koch AD, et al. | Colonoscopy, both groups received VR training but different amounts | 18 (n=9, i=10) i=30 total, i=100 total | Yes: to assess the clinical performance of novice endoscopists during colonoscopy after intensive and prolonged training on a VR endoscopy simulator | Can't tell |
| 83    | 14                 | Zandaloo B, et al. | Laparoscopic inguinal hernia repair | 50 (n=24, i=26, crossover from control to Intervention after TEP)=10 | Yes: to evaluate a mastery learning, simulation based curriculum for laparoscopic, TEP inguinal hernia repair | Can't tell |
| 85    | 15                 | Kassler DO, et al. | Lumbar puncture on infants, self reported success after simulation training | 56 (n=28, i=28) reported clinical LP (n=17, i=15) completed final OSCE assessment (n=29, i=26) | Yes: to demonstrate that deliberate practice simulation based training improves infant LP skills compared to a control group | Can't tell |
| 91    | 16                 | Ceballosfe D, et al. | Warm-up in simulator or not before laparoscopic cholecystectomy | 10 (8 actually analyzed in results) | Yes: to determine if a short VR warm-up curriculum improves laparoscopic performance in the OR | Can't tell |
| 93    | 17                 | Heynich A, et al. | Simulator training in colonoscopy | 40 (n=20, i=20) included in final analysis (n=18, i=18) | Yes: the educational evaluation of the simulator as a tool for training novices in colonoscopy | Can't tell |
| 101   | 18                 | Ahlborg G, et al. | VR training in cholecystectomy | 13 (n=6, i=7) | Yes: to assess the effect of proficiency-based VR training on the outcome of the first 10 entire cholecystectomies performed by novices | Can't tell |
| 104   | 19                 | Cohen J, et al. | Simulator training in colonoscopy | 51, completed study 45 (n=23, i=22) | Yes: to define the benefit of training on the GI mentor on competency acquisition in colonoscopy | Can't tell |

*See Data Sheet (Supplemental File 4) for treatment effect sizes and precision of estimated treatment effects
Columns F - P correspond to CASP Randomized Controlled Trial Checklist items 1 - 11.
| Were all of the patients who entered the trial properly accounted for at its conclusion? | Were patients/health workers/study personnel 'blind' to treatment? | Were the groups similar at the start of the trial? | Aside from experimental intervention, were the groups treated equally? | How large was the treatment effect? | How precise was the estimate of the treatment effect? |
|---|---|---|---|---|---|
| Can't tell | Authors claim observers were blinded, no mention of whether or not patients were blinded | Can't tell | Yes | Can't tell | Can't tell |
| Yes | Non-blinded observer recorded operative metrics and consultant tookovers. Blinded consultant assessed GRSS and examiner checked after procedure. A blinded observer also assessed these post-hoc based on fluoroscopy footage and hand movements. | Yes, with respect to sex, post-grad year and number of endovascular cases assisted | Yes, but post testing of the control group was completed | Can't tell | Can't tell |
| Can't tell | Porcine and OR performance assessed using BOSATS scale by 1 trained and blinded rater. Non-technical skills assessed by NOSTS system by 2 trained raters, 1 blinded and 1 non-blinded. | Yes, with respect to a host of variables, but intervention group had significantly fewer basic bariatric surgeries performed as the primary surgeon and bariatric rotations participated in | Yes, but the control group did perform the surgery being assessed on a human in the OR | Can't tell | Can't tell |
| Can't tell | Can't tell if patients were blinded. Researchers assessing and enrolling patients, as well as outcome assessors were blinded. | Yes, inclusion criteria were age>18, non-rutputured aortic aneurism, suited to treatment with Gire Excluder AAA endoprosthesis/Endurant stent graft | Can't tell | Can't tell | Can't tell |
| Can't tell | Can't tell if patients were blinded. | | | Can't tell | Can't tell |
| Can't tell | Assessors were independent and blinded. Can’t tell if patients were blind to treatment | Yes with respect to age, sex, post-grad year and arthroscopies performed pre and post intervention | | Yes | Can’t tell |
| Can’t tell | Assessors of operative performance were blinded. Assessment of pre-operative non-technical skills was not blinded. Not clear if assessment on box trainer and simulator was blinded. | Yes with respect to a host of variables (surgical experience, VR experience, musical instrument experience etc.) | | Yes | Can’t tell |
| Can’t tell | Yes, single OSAT evaluator was blinded | Groups stratified by pre-intervention human salpingectomy OSAT score, post grad year was similar | | Yes | Yes |
| Can’t tell | ASSET score assessed by 2 blinded evaluations. Can’t tell if patients were blind to treatment | Yes with respect to sex, post-grad year and number of cases performed | | Yes | Yes |
| Can’t tell | Supervisor (resident, fellow or attending) and the patients nurse rated the CVC placements and were blinded | Yes, with respect to age, sex, training track and degree | | Yes | Yes |
| Can’t tell | JAG DOPS, Global Rating Form and Communication GRS were all blinded. Can’t tell if patients were blind to treatment. | | | Yes | Yes |
| Can’t tell | Assessors of surgical time and technical score were blinded | Yes, with respect to age, sex, time in surgical employment and prior number of performed hernia repairs | | Yes, but control group was assessed 2 times compared to the intervention group which was assessed 4 times | Yes |
| Can’t tell | Supervising surgeon was blinded to amount of VR simulated procedures the trainee had completed at the time of patient based assessment | Yes, all subjects were at the start of their training in gastroenterology with no prior endoscopic experience | Yes, but intervention group i completed 90 total VR colonoscopies and group ii completed 100 total VR colonoscopies | Can’t tell | Can’t tell |
| Can’t tell | Supervising staff surgeon was blind to the status of the resident surgeon, observer was not blinded (Staff GOALS score was used, both reviewed video recording to assess intraoperative complications). Retrospective assessment of patient medical records was done by a blinded member of staff. | Yes, baseline TEP repair was similar, groups were similar with respect to a host of other variables (post-grad year, sex, handedness, video game experience, TEP comfort + experience) | | Yes, but the control group could cross over to the intervention protocol after TEP2, 10 participants elected to do this | Can’t tell |
| Can’t tell | OSCE checklist assessors were not blinded, but 69% of the tests were assessed post hoc (video recording) by a blinded author. | Yes, with respect to sex, post-grad year and experience with LP (training, simulator experience, observations, LP performed) | | Yes | Yes |
| Can’t tell | OSATS score given by 2 blinded raters | Each surgeon served as their own control | | Yes | Yes |
| Can’t tell | Expert assessors of patient colonoscopies were blinded | Yes, with respect to age, sex, educational direction, sigmoidoscopies and colonoscopies witnessed/assisted/perform | | Yes | Yes |
| Can’t tell | Surgical supervisors were blinded. The 2 video assessors were blinded. | Yes, with respect to age, sex, visuospatial assessment, working memory assessment and laparoscopic assisting experience | | Yes | Yes |
| Can’t tell | Proctors who assessed fellows after the procedure were blinded | Yes, with respect to experience with gastroscopy and flexible sigmoidoscopies | | Yes | Yes |
### Measured outcomes: operative time, contrast volume, fluoroscopy time, time to carotid cannulation and carotid sheath duration

**Cost effectiveness not assessed**

**No significant effect demonstrated**

**ANGIO Mentor; Simbionics, Aport City, Israel**

**Time until proficiency; defined as attaining an MCQ and passing technical proficiency on simulator 2 times at all 4 stages of the program. Mean time to completion was 6.44 months**

**Cost effectiveness not assessed**

**No significant effect demonstrated**

**LapSim; Surgical Science, Gothenburg, Sweden**

**Seven 2 hour sessions on box trainer and VR simulator. Access to these was not restricted to either group outside of the mandatory sessions. 1 session for intervention group on non-technical skills 10-15min.**

**Cost effectiveness not assessed**

**No significant effect demonstrated**

**OSAT score**

**Cost effectiveness not assessed**

**Porcine cadavers**

**1 hour to familiarise, 0.5 hours after to do 1 surgery**

**Cost effectiveness not assessed**

**No significant effect demonstrated**

**Artho VR shoulder simulator; Simbionics, Cleveland, Ohio, USA**

**1 hour**

**Cost effectiveness not assessed**

**Venous access simulator, Blue Phantom, Redmond, Washington, USA**

**1.2 hours of individual training, then training as needed to pass independent CVC placement**

**Cost effectiveness not assessed**

**EndoVR, CAE Healthcare Canada, Quebec, Montreal, Canada**

**6 hours of practice on simulators (control did not get instruction or feedback from professional)**

**Cost effectiveness not assessed**

**"plastic phantom mimicking the human male groin" and surgery of anesthetized pig with congenital hernia**

**1 day skills lab course with practive on plastic phantom and anaesthetized pig**

**Cost effectiveness not assessed**

**GI Mentor II, Simbionics, Cleveland, Ohio, USA**

**Amount of completed VR cases; group i was tested twice in patient after 10, 30 and 50 VR cases. Group ii was tested twice in patients after 20, 60 and 100 VR cases completed.**

**Cost effectiveness not assessed**

**GI Mentor, Simbionics, Cleveland, Ohio, USA**

**Simulation practice until mastery was achieved, defined as successful repair of both hernias (bilaterally) in less than 3 minutes on 2 consecutive attempts.**

**Cost effectiveness not assessed**

**Baby Step neonatal task trainer; Laerdal, Wappinger Falls, New York, USA**

**Simulation practice until mastery was achieved, defined as demonstrating all steps (15 point checklist) flawlessly and independently from start to finish on the simulator.**

**Cost effectiveness not assessed**

**LapSim; Surgical Science, Gothenburg, Sweden**

**3 laparoscopic tasks (7 available at 3 different difficulty levels), lasting approx. 15min**

**Cost effectiveness not assessed**

**Endo TS-1; Olympus Keymed, Southend, UK**

**Intervention group received 16 hours of practice on simulator. Control group received 16 hours of practice on patients (4 half day sessions, required to perform a minimum of 8 procedures per session)**

**Cost effectiveness not assessed**

**LapSim; Surgical Science, Gothenburg, Sweden**

**Intervention group received simulator training until they achieved proficiency, defined as achieving the median score of 5 expert laparoscopic scores 6 task twice**

**Cost effectiveness not assessed**

**GI Mentor, Simbionics, Cleveland, Ohio, USA**

**Intervention group practiced on the VR simulator for 10 hours prior to performing live colonoscopies**
Control group training | Which patient outcomes were recorded? | Further notes
---|---|---
None | None | None
Continued conventional training | Recorded patient outcomes: peri-operative complications, major or minor adverse events in hospital and 30 days after treatment | 16 hours of practice on patients and controls. Surgeons served as their own controls. There were no differences in peri-operative complications, major or minor adverse events in hospital and 30 days after treatment between intervention and control at several points in time. Residents become competent at the same rate. The intervention group was “ahead on the learning curve” compared to the control group. There was a significant improvement in technical + clinical success rates, in-hospital + 30 day mortality and in the retention, recurrence of hernia and groin pain 3-months retention, post-post-op complications (hematoma, seroma, skin infection and urinary tear, conversion of surgical approach (open or trans-abdominal), post-operative complications (infections, anemia, skin-infection and urinary retention), post-op complications (overnight stay, recurrence of hernia and groin pain 3-months post repair). Subjective competence was significantly better in the intervention group until they achieved mastery. At TEP3 the intervention group was better, completed more of the procedure themselves and had lower rates of complications. At TEP5 the residents crossed over from control to intervention performed better than their control counterparts. At TEP3/4/5 GOALS scores were not significantly different between the groups. Complications were similar between crossover and control groups at TEP3. All TEPs after intervention combined, and excluding the crossover group, the intervention group were statistically better in all measured outcomes (aside from overnight stay). Seem like a slam dunk. Why is crossover and control similar at TEP3, does that mean intervention and control are similar at TEP3? Difference in total measures is significant but groups compared at TEP3/4/5 aren’t.

Rehearsal after procedure | Recorded patient outcomes: peri-operative errors, technical + clinical success rates, in-hospital + 30 day mortality | Surgery performed better on sim, sim test before intervention, right after then 4-6 weeks after. Intervention group was statistically better than control group with respect to JAG DOPS score at clinical colonoscopies. No patient outcomes measured. Highlights differences between running sim practice alone and getting feedback from expert.

Sim training without feedback | None | None
Continued conventional training | Recorded patient outcomes: arterial puncture, hematoma, catheter malposition, catheter associated infection, pneumothorax and death | Intervention group received 16 hours of sim practice, control group received 16 hours of practice on patients. At post intervention assessment the intervention group outperformed control on the simulator. There were no statistically significant differences between groups upon patient based assessment. Interesting because it compares patient and sim practice directly.

Simulation training without feedback | None | None
Continued conventional training | None | None
Continued conventional training | Recorded patient outcomes: retroperitoneal complications (wound + blader injury, perforated tear, conversion of surgical approach (open or trans-abdominal), post-operative complications (infections, anemia, skin-infection and urinary retention), post-op complications (overnight stay, recurrence of hernia and groin pain 3-months post repair) | Intervention group trained on simulator until they achieved mastery. At TEPP the intervention group was better, completed more of the procedure themselves and had lower rates of complications. All TEPs the residents crossed over from control to intervention performed better than their control counterparts. At TEPP/4/5 GOALS scores were not significantly different between the groups. Complications were similar between crossover and control groups at TEPP. All TEPs after intervention combined, and excluding the crossover group, the intervention group were statistically better in all measured outcomes (aside from overnight stay). Seem like a slam dunk. Why is crossover and control similar at TEPP, does that mean intervention and control are similar at TEPP? Difference in total measures is significant but groups compared at TEPP/4/5 aren’t.

Surgeons served as their own controls | None | None
Continued conventional training | 16 hours of practice on patients | 8 surgeons completed 2 surgeries each, 1 with warm up before and 1 without. Surgeons with warm-up beforehand received significantly better OSATS scores. No other variables measured. Small sample size. Intervention group received 16 hours of sim practice, control group received 16 hours of practice on patients. At post intervention assessment the intervention group outperformed control on the simulator. There were no statistically significant differences between groups upon patient based assessment. Interesting because it compares patient and sim practice directly.

Continued conventional training | None | None
Continued conventional training | None | None
Continued conventional training | None | None
Continued conventional training | None | None
Continued conventional training | None | None
Continued conventional training | None | None