Author, year, title	Geography	Disciplines	Skill studied	Learner number and type	Type of simulation modality	Simulation instructional design variation studied	Kirkpatrick level	Outcome
Chunharas et al. (2013) Medical students themselves as surrogate patients increased satisfaction, confidence, and performance in practicing injection skill	Thailand	Paediatrics	Subcutaneous and intramuscular injection	89 medical students (5th year)	Manikin (model unspecified) and fellow students	Manikin (model unspecified) vs Manikin (model unspecified) and fellow students	3	Observed performance in injections performed on real children, using a rating scale with minimal validity evidence described. Improved performance in the manikin + fellow student group for preparing children and giving the injection. No difference in other steps including checking accuracy of order, preparing vaccine, selection of injection site, sterile technique, handling of syringe and needle, filling medical record, and explaining purpose and effect of vaccine.
Daly et al. (2013) Efficacy of surgical simulator training versus traditional wet-lab training on operating room performance of ophthalmology residents during the capsulorhexis in cataract surgery	USA	Ophthalmology	Capsulorhexis	21 ophthalmology residents (2nd year)	Eyesi virtual reality simulator and silicone eyes in a wet lab with the same equipment as an actual operating room	Eyesi virtual reality simulator vs Silicone eyes in a wet lab with the same equipment as an actual operating room	3	No difference in an overall score consisting of both process (economy of movement, confidence of movement, errors in tissue handling) and product variables (time, size, continuity, shape and centring of capsulorhexis). Minimal validity evidence described for rating tool used.
de Melo et al. (2018) Self-perceived long-term transfer of learning after postpartum haemorrhage simulation training	Brazil	Obstetrics and Gynaecology	Post-partum haemorrhage management	12 residents	Part task pelvis simulator (ProDelphus) with simulated patient and simulated nurse	Simulation aligned to a complex bundle of instructional design principles: authenticity, psychological fidelity, engineering fidelity, Paivio's dual coding, feedback, variability, increasing complexity	3	During individual interviews, participants reported improved transfer of communication and teamwork skills and situational awareness in the clinical environment.

						vs Simulation designed according to existing practice		
De Win et al. (2016) An evidence-based laparoscopic simulation curriculum shortens the clinical learning curve and reduces surgical adverse events	Belgium	General Surgery	Laparoscopic cholecystic-tomy	30 medical students (final year) who transited into surgical residency	Progression through multiple simulators including suture pad, chicken skin, pulsatile organ perfusion trainer, living rabbit model	Simulation training with proficiency-based progression vs Simulation training without proficiency-based progression	4b	Lower odds of adverse events (bleeding or liver damage) with proficiency-based progression during laparoscopic cholecystectomy on actual patients.
DeStephano et al. (2015) A randomized controlled trial of birth simulation for medical students	USA	Obstetrics and Gynaecology	Vaginal delivery	110 medical students	Birth simulator manikins	Noelle: High cost, high technological complexity, low portability, standalone vs MamaNatalie: Low cost, low technological complexity, high portability, hybrid simulation (worn on patient)	3	No difference in performance of vaginal delivery steps as rated by preceptors using a previously established checklist.
Dyre et al. (2017) Imperfect practice makes perfect: error management training improves transfer of learning	Denmark	Obstetrics and Gynaecology	Obstetric ultrasound including foetal weight estimation	60 medical students (5^{th} and 6^{th} year)	Transabdominal ScanTrainer	Error management training vs Error avoidance training	3	Higher scores on the OSAUS scale in EMT group (67.7%) than EAT group (51.7%) when assessing foetal weight in actual pregnant patients. Deviation in foetal weight estimated by participant from weight estimated by expert was 16.7% in EMT group and 26.6% in EAT group, but this difference was not statistically significant.

Egenberg et al. (2017) "No patient should die of PPH just for the lack of training!" Experiences from multi-professional simulation training on postpartum haemorrhage in northern Tanzania: a qualitative study	Tanzania	Obstetrics and Gynaecology	Post-partum haemorrhage prevention, management and communica-tion	42 Midwives, medical attendants, doctors	Multiprofessional simulation training, with technical skills training on MamaNatalie	Qualitative study using focus group discussions, in the context of a related study experienced by the participants, that examined multiprofessional training for post-partum haemorrhage.	3	Improved communication, leadership, and clinical management of post-partum haemorrhage.
Gomez et al. (2015) Evaluation of two flexible colonoscopy simulators and transfer of skills into clinical practice	USA	Surgery	Colonoscopy	27 surgical residents (PGY-1)	Endoscopic virtual reality (GI Mentor II) and physical model simulators (Kyoto Kagaku)	Endoscopic virtual reality vs Physical model simulators vs Both	3	Improvement seen from pre-test to post-test in the groups that used GI mentor alone or both simulators, compared to physical model alone, as measured by the GAGES-C tool when performing colonoscopy on a real patient.
Grover et al. (2017) Progressive learning in endoscopy simulation training improves clinical performance: a blinded randomized trial	Canada	Internal Medicine	Colonoscopy	37 residents with <20 previous endoscopies	Bench top simulator (physical) and EndoVR (virtual reality) endoscopy simulator	Progressive task difficulty (1 hour bench top then 5 hours EndoVR cases in increasing difficulty) vs Random order of task difficulty (6 hours of EndoVR with random order of task difficulty)	3	Progressive group outperformed the random order group as measured by the JAG DOPS tool during colonoscopies on real patients.
Hernández-Irizarry (2016) Optimizing training cost-effectiveness of simulation-based laparoscopic inguinal hernia repairs	USA	General Surgery	Laparoscopic inguinal hernia repairs (deemed as a high complexity, low organization task)	44 residents (PGY-1 to 5)	Guildford MATTU TEP task trainer, an inanimate box trainer	Randomised part task vs Whole task	3	Participants in the part task group achieved mastery of the skills curriculum on average 17 minutes quicker than those in the whole task group (60 vs 77 mins), with no difference in GOALS scores when performing surgeries in actual patients.

Kessler et al. (2015) Impact of just-in-time and just-in-place simulation on intern success with infant lumbar puncture	USA	Paediatrics	Infant lumbar puncture	1319 interns who performed 436 infant lumbar punctures	Infant lumbar puncture physical simulator (BabyStap)	Simulation based training to mastery standard plus just in time and just in place training vs Simulation based training to mastery standard alone	4b	No significant differences in first infant lumbar puncture success rate. JIT group had lower mean number of attempts (1.4 vs 2.1), and increased use of early stylet removal, analgesia, and family presence.
Kroft et al. (2017) Preoperative practice paired with instructor feedback may not improve obstetrics-gynaecology residents' operative performance	Canada	Obstetrics and Gynaecology	Laparoscopic salpingectomy	18 PGY-2 to 6 trainees	LapSim virtual reality surgical simulator	Preoperative practice with feedback, with feedback based on preoperative practice within one hour before surgery vs Preoperative practice alone vs Feedback alone based on baseline testing	3	No significant difference as measured by objective structured assessment of laparoscopic salpingectomy, performed on real patients.
Lal et al. (2022) Evaluating the optimal training paradigm for transcarotid artery revascularization based on worldwide experience	Countries not specified in manuscript	Vascular Surgery, Neurosurgery, Interventional Cardiology, Interventional Radiology, Cardiothoracic Surgery	Transcarotid-artery revascularize-tion	1160 physicians credentialed to perform carotid endarterectomy at home institution	Human cadavers and synthetic models	Supervised training on human cadavers vs Supervised training on synthetic models	4b	No difference in rates of clinical adverse outcomes or technical adverse events.

Liao et al. (2013) Coached practice using ERCP mechanical simulator improves trainees' ERCP performance: a randomized controlled trial	Taiwan	Gastroentero-logy	Endoscopic retrograde cholangiopanc reatography	16 fellows	Mechanical simulator	Coached practice (6 hours) followed by uncoached practice (1 hour every 2 weeks for 3 months) in 2009 vs Coached practice (6 hours) followed by no further simulation training in 2008 vs No simulation training in 2008 and 2009	3	Coached + uncoached was indirectly compared with Coached alone, in that both of these were first compared to control groups in their respective study years, and adjusted odds ratios were then compared for successful deep biliary cannulation in real patients, with no difference shown.
Mankute et al. (2022) A novel algorithm-driven hybrid simulation learning method to improve acquisition of endotracheal intubation skills: a randomized controlled study	Lithuania	Emergency Medicine and Anaesthesia	Endotracheal intubation	77 medical students (5^{th} year) and residents (1^{st} year)	Manikin with teacher (3 hours) then without teacher (3 hours) Manikin with handheld tablets containing scenarios, checklists, and learning algorithms (6 hours)	Lectures, then manikin with teacher (3 hours) then without teacher (3 hours) in groups of 10 vs Review lectures, videos and algorithms on a virtual learning environment, followed by 10 question pre-test, then manikin with handheld tablets containing scenarios, checklists, and learning algorithms (6 hours at learners' own pace, with 24/7 access to simulation lab) in groups of 3 for peer-to-peer practice, with remote review of practice video recordings by teachers and feedback by email	3	More learners performed more actions correctly, and in the correct sequence, as assessed by a checklist with minimal validity evidence, on actual patients.

Mduma et al. (2015) Frequent brief on-site simulation training and reduction in 24-h neonatal mortality—an educational intervention study	Tanzania	Neonatology, Obstetrics and Gynaecology	Delivery room management of new born	Unclear number of midwives, nurse students, operating nurses, and doctors. 9708 deliveries were studied.	NeoNatalie	Frequent and brief on-site simulation (40 mins a month + 3 mins a week) vs One-day simulation course	4b	Reduced infant death within 24 hours of birth (11.1/1000 vs 7.2/1000). More neonates were stimulated (14.5% vs 16.3%).
Naples et al. (2022) The impact of simulation training on operative performance in general surgery: lessons learned from a prospective randomized trial	USA	General Surgery	Bowel anastomosis	9 interns	Porcine intestine	Proficiency based training: needed a perfect score at end of simulation assessment prior to completing a post-test. If not, practice independently and re-assess vs No required score at end of simulation assessment before proceeding to post-test	3	No difference between groups in operative performance with actual patients as measured by ACS/APDS global rating scale.
Nilsson et al. (2017) Simulation-based camera navigation training in laparoscopy-a randomized trial	Denmark	General Surgery and Gynaecology	Camera navigation during laparoscopic surgery	36 medical students (4^{th} to 6^{th} year)	LapSim virtual reality surgical simulator	Simulation based part task practice (camera navigation) vs Simulation based whole task practice (cholecystectomy)	3	No significant difference in camera navigation skills (part task) during cholecystectomies performed on actual patients, as measured by the authors' own newly derived tool with minimal validity evidence, the objective structured assessment of camera navigation skills.

O'Sullivan et al. (2014) The effect of simulation-based training on initial performance of ultrasound-guided axillary brachial plexus blockade in a clinical setting - a pilot study	Ireland	Anaesthesia	Ultrasound guided axillary brachial plexus blockade	10 residents with no prior experience with ultrasound guided regional anaesthesia	Hands on simulation with cadavers, ultrasound scanning of a volunteer, needling skills sessions with tissue phantoms, and a novel simulator (PHANTOM Desktop device) with physical devices to manipulate, haptic feedback, and a computer monitor with 3D glasses to visualise virtual actions	Control: Hands on simulation with cadavers, ultrasound scanning of a volunteer, needling skills sessions with tissue phantoms vs Intervention: practice as per control group, plus additional practice by completing 4 tasks to a predefined proficiency level on PHANTOM Desktop device, which provided computer generated feedback	3	No difference between groups in performance as measured by the sum of scores on a global rating scale and checklist, from observer ratings of participants performing ultrasound guided axillary brachial plexus blockade on real patients.
Orzech et al. (2012) A comparison of 2 ex vivo training curricula for advanced laparoscopic skills: a randomized controlled trial	Canada	General Surgery	Laparoscopic suturing (as an advanced rather than basic laparoscopic skill)	24 residents PGY-2 or above with >10 prior laparoscopic procedures	LapSim virtual reality simulator and Fundamentals of Laparoscopic Surgery box trainer	LapSim virtual reality simulator with 3 progressive settings of difficulty, with no supervision by surgeons vs Fundamentals of Laparoscopic Surgery box trainer with supervision by surgeons In both groups, practice was carried out until proficiency criteria were reached	3	No differences between VR and box trainer groups as measured by performance in placing intracorporal laparoscopic stitches during a Nissen fundoplication on a real patient, using a procedure-specific checklist and global rating scale.

Patel et al. (2012) Operating room introduction for the novice	UK	General Surgery, Gynaecology and Otolaryngology	Appropriate behaviour for initial attendance within operating room	60 medical students (1st year)	Second Life online operating room vs Physical simulated operating room	Second Life online operating room vs Physical simulated operating room	3	Physical simulated operating group performed better than the Second Life online operating room group, as measured by a checklist observation scale with minimal validity evidence described, while students were in actual operating rooms.
Schaffer et al. (2021) Association of simulation training with rates of medical malpractice claims among obstetrician-gynaecologists	USA	Obstetrics and Gynaecology	Management of obstetric emergencies	292 attending obstetricians and gynaecologists	Not specified. Team training and crisis management, rather than surgical or technical skills	Single simulation session vs More simulation sessions	4a	Attending more simulation sessions was associated with a reduced malpractice claim rate.
Sharara-Chami et al. (2014) Simulation training in endotracheal intubation in a paediatric residency	Lebanon	Paediatrics	Endotracheal intubation	10 residents	Manikin (SimBaby, Laerdal)	SimBaby manikin with vital signs displayed on monitor vs SimBaby manikin with vital signs and physical examination findings read out by supervisor	4a	No difference in number of successful intubations logged by participant.
Shore et al. (2016) Validating a standardized laparoscopy curriculum for gynaecology residents: a randomized controlled trial	Canada	Obstetrics and Gynaecology	Laparoscopic surgery	27 residents (PGY-1 to 2)	Box trainer, VR simulator, SimMan physical patient simulator	Structured simulation group with cognitive training (lectures, videos, reading materials), box trainer for technical skills with pre-set proficiency benchmarks, VR simulator for technical skills, and non-technical skills training with scripted confederates and SimMan physical patient simulator with	3	Structured simulation curriculum group performed better than the conventional curriculum group as measured by OSA-LS when performing laparoscopic right salpingectomy and intracorporal knot tying of the left round ligament on a real patient.

						debriefing. Total time: 3 hours a week for 7 weeks vs Conventional curriculum, part of which includes simulation without further details		
Sørensen et al. (2015) Clarifying the learning experiences of healthcare professionals with in situ and off-site simulation-based medical education: a qualitative study	Denmark	Obstetrics and Gynaecology, Anaesthesia	Management of obstetric emergencies	25 obstetricians, midwives, auxiliary nurses, anaesthetists, nurse anaesthetists, operating room nurse	In situ simulation and Off-site simulation	In situ simulation vs Off-site simulation	4a	Focus group discussion themes: - Participants perceived ISS and OSS had the same effect on individual and team learning - ISS had more organizational impact and provided more information for practical organizational changes - Physical context and physical fidelity were not the most important, provided that psychological and sociological authenticity elements are respected, such as participant preferences for simulation in one's own authentic role - OSS had the positive effect of forcing participants to adapt to new places and people and forced them to see their own routines from the outside - Perceptions about ISS and OSS differed between professional groups, with nurses preferring equipment to be in the right place (thus ISS).

Srinivasan et al. (2018) Proficiency-based progression training: an 'end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomized control study	Ireland	Anaesthesia	Epidural analgesia during labour	17 residents with <2 years of experience and <50 prior epidural catheter placements	Manikin KKM43E, Cardiac services 2013, SISK Healthcare Group	Proficiency based progression simulation training vs Simulation training without proficiency-based progression	4b	Proficiency based progression group had fewer epidural failures (13.3%) than the simulation group without proficiency-based progression (28.7%).
Tan et al. (2018) Teaching residents chest tubes: simulation task trainer or cadaver model?	USA	Emergency Medicine and General Surgery	Chest tube insertion	16 residents (PGY-1 to 2)	Manikin (TraumaMan) vs Cadaver	Manikin (TraumaMan) vs Cadaver	4a	No difference between groups in ability to insert a chest tube independently (i.e. without supervisor assistance) at first attempt in a real patient, as self-reported through case logs.
Tchorz et al. (2015) Pre-clinical endodontic training with artificial instead of extracted human teeth: does the type of exercise have an influence on clinical endodontic outcomes?	Germany	Dentistry	Root canal	89 dentistry students (3rd year)	Artificial resin teeth and extracted human teeth	Artificial resin teeth vs Extracted human teeth	3	No significant difference between groups in proportion of patients with acceptable outcomes, based on review of post procedural radiographs.
Todsen et al. (2013) Short- and long-term transfer of urethral catheterization skills from simulation training to performance on patients	Denmark	Not specified	Urethral catheterization	64 medical students (3rd year) with no prior urethral catheterization experience	Male manikin with an actor sitting behind manikin	Simulation, followed by just-in-time video before performance on real patients vs Simulation without just-in-time video before performance on real patients	3	No significant difference as measured by a checklist when performing urethral catheterization on a real patient.

Tolsgaard et al. (2015) The effect of dyad versus individual simulation-based ultrasound training on skills transfer	Denmark	Obstetrics and Gynaecology	Transvaginal ultrasound	30 final year medical students with minimal prior ultrasound experience	ScanTrainer transvaginal ultrasound simulator: physical device with similar shape to a transvaginal probe with haptic feedback and monitor output with images obtained from real patients. But not a manikin. 9 modules over 2 hours.	Dyad learning: each individual attempted every module once. vs Individual: each individual attempted every module twice.	3	No significant difference in mean scores between dyad (56.3) and individual (48.4) groups on the OSAUS on transvaginal ultrasound scans performed on real patients. A larger proportion of the dyad group (71%) scored above the pre-established criterion referenced pass fail level than the individual group (30%).
Walsh et al. (2020) Non-technical skills curriculum incorporating simulation-based training improves performance in colonoscopy among novice endoscopists: Randomized controlled trial	Canada	Gastroentero-logy, General Surgery and Internal Medicine	Non-technical skills for colonoscopy	39 postgraduate trainees who completed <20 endoscopies	Bench top model and EndoVR model	Colonoscopy simulation training with expert feedback vs Colonoscopy simulation training with expert feedback. Expert focused on NTS with access to NoSEy checklist for NTS	3	NTS group performed better as measured by the JAG DOPS tool during colonoscopies of actual patients. JAG DOPS measures both technical skills and NTS. It is unclear from the manuscript whether the difference was driven purely by NTS, although at level 2 outcomes, both NTS and technical skills were better.

Abbreviations. ACS/APDS: American College of Surgeons / Association of Program Directors in Surgery; EAT: Error avoidance training; EMT: Error management training; GAGES-C: Global Assessment of Gastrointestinal Endoscopic Skills-Colonoscopy; GOALS: Global Operative Assessment of Laparoscopic Skills; ISS: In situ simulation; JAG DOPS: Joint Advisory Group Direct Observation of Procedural Skills; JIT: Just in time; OSA-LS: objective structured assessment of laparoscopic salpingectomy; NTS: Non-technical skills. OSAUS: objective structured assessment of ultrasound; OSS: Off-site simulation; PGY: Post graduate year; UK: United Kingdom; USA: United States of America; VR: Virtual reality.

Kirkpatrick levels. 1: Reaction e.g. participants' views on learning experience; 2a: Learning – Change in attitudes; 2b: Learning – Modification of knowledge or skills; 3: Behaviour – Change in behaviours; 4a: Results – Change in the system/organisational practice; 4b: Results – Change in patient outcomes.

Table 1. List of included studies and skills, instructional design variations and outcomes examined