Low versus high level of physical resemblance in simulation for the acquisition of basic surgical skill: a meta-analysis

Fabrizio Consorti, Gianmarco Panzera

ABSTRACT

Background Many studies explored the use of simulation in basic surgical education, with a variety of devices, contexts and outcomes, with sometimes contradictory results.

Objectives The objectives of this meta-analysis were to focus the effect that the level of physical resemblance in a simulation has on the development of basic surgical skill in undergraduate medical students and to provide a foundation for the design and implementation of a simulation, with respect to its effectiveness and alignment with the learning outcomes.

Study selection We searched PubMed and Scopus database for comparative randomised studies between simulations with a different level of resemblance. The result was synthesised as the standardised mean difference, under a random effect model.

Findings We selected 12 out of 2091 retrieved studies, reporting on 373 undergraduate students (mean of subjects 15.54±6.89). The outcomes were the performance of simple skills and the time to complete a task. Two studies reported a scoring system; seven studies reported time for a task; and three studies reported both. The total number of measures included in the meta-analysis was 456 for score and 504 for time. The pooled effect size did not show any significant advantage in a simulation of a high level of physical resemblance over a lower level, both for the scoring system (−0.19, 95% CI −0.44 to 0.06) and for time (−0.14, 95% CI −0.54 to 0.27).

Conclusion Simulations with a low level of physical resemblance showed the same effect as the simulation using a higher level of resemblance on the development of basic surgical skills in undergraduate students.

INTRODUCTION

“The earlier surgical skills training starts the better”. This was one of the 12 tips Kneebone listed more than 20 years ago, highlighting the importance of an early beginning of surgical training in a medical curriculum. Actually, many international curricula and core competence inventories include basic surgical skill as a component of professional competence. Clerkship remains the cornerstone of surgical undergraduate education, and many authors focused on the level of fidelity in various clinical domains. Hamstra et al questioned the term ‘fidelity’ and argued that the concept of fidelity is multidimensional. They suggested using the terms physical resemblance, to denote the physical structure of the device used in simulation, and functional task alignment, to denote the coherence of the whole context of simulation with the expected learning outcomes. Finally, they recommended giving more emphasis to the transfer of learning, that is, the ability to apply the skill trained in a context to a different context.

There are theoretical reasons—discussed in the next section—to hypothesise that the learning outcomes included in the class of basic surgical skill can be achieved with a low level of physical resemblance. The objective of this meta-analysis was to review and synthesise the available evidence about the effect that the level of physical resemblance in a simulation has on the development of surgical skill in undergraduate students, despite many studies exploring the use of simulation in this field, with a variety of devices, context and outcomes.

The design of this meta-analysis has been registered in the PROSPERO database (CRD42019128061).

Theoretical background

In this section, first, we give a definition of basic surgical skill and of undergraduate student, then we consider some of the educational theories that can predict the role of the characteristics of simulation in its effectiveness and we briefly discuss them.

We considered as basic surgical skills both simple tasks like gowning and gloving, suturing, knot tying and a set of psychomotor abilities like dexterity and coordination of the two hands, or the ability to work with a 2-D view. This last set of skills is important because of the increasing importance that minimally invasive surgery has. Among the others, many authors focused on the level of fidelity in various clinical domains.
Behaviourism is one of the most frequent theoretical positions to frame simulation. It posits that the outcome of learning is the modification of behaviour, regardless of the implicated cognitive mechanisms. Teaching is then providing an adequate stimulus to obtain the desired response. An effective training starts from easy tasks and progresses through an intentional sequence up to proficiency. Deliberate practice and mastery learning are typical behaviourist instructional methods. Also, the use of a part-task trainer in simulation is justified in a behaviourist approach.

Cognitivism is a class of theories that explore the mental mechanisms in learning. In this class, situated cognition theory maintains that learning emerges from the interplay of the individual cognitive process and the social, cultural and physical environment in which the process takes place. This theory justifies the effectiveness of simulations with a high level of physical resemblance, in which the action takes place in an environment that recalls the real professional situation. Another relevant contribution of the cognitivist perspective is the concept of mental workload. Our mind has a limited capacity of working memory, and an overload of cognitive and sensory information can alter mental performances like learning or taking clinical decisions. Naismith and Cavalcanti reviewed the effect of cognitive load in simulation, showing that an excess of load impairs learning.

The learning model known as the challenge point framework builds on both behaviourist and a cognitivist perspectives. This model describes the relationship between the performance and the difficulty of a task, defined as the physical or cognitive challenge posed by a motor problem. The model proposes a progression from simpler to more difficult tasks as the most effective strategy for learning.

Based on these premises, if basic surgical skills are the building blocks of a specific professional competence to be further developed in the future of graduate education, then we hypothesise that a simulation aligned to these basic outcomes and based on a low level of physical resemblance is as effective as simulation with a higher level of resemblance and that reproduces a realistic professional context.

**MATERIALS AND METHODS**

We planned, conducted and reported this study according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards of quality.

**Study eligibility**

This meta-analysis considered trials in which undergraduate students were randomised to use different equipment for the simulation of surgical skills. For laparoscopic simulation, we considered a simulation in which either immersive virtual reality or 3-D vision was used as having a higher level of physical resemblance than the simulations in which a box trainer was used, either with direct vision or based on a conventional video system. For non-laparoscopic simulation, a high physical resemblance was defined as the use of a biological model as opposed to the use of a simplified synthetic model. We excluded studies about endoscopic skill (gastroscopy and bronchoscopy) because these skills did not fit with the definition we gave of basic surgical skills. Studies comparing simulation with a no-intervention or no-simulation group were excluded. In studies with three arms, we considered only the two arms comparing devices that we interpreted as high or low resemblance. The articles had to provide the figures needed to compute the overall effect size, and the considered outcome measures were both any scoring system for skill and time for a task. We selected articles in any language, without time limits.
**Systematic review**

**Table 1  Characteristics of the selected studies**

| Article          | Subjects | Skill                          | Comparison              | Outcome           | Measure            | Result             |
|------------------|----------|--------------------------------|-------------------------|-------------------|--------------------|--------------------|
| Brinkman et al.  | 18 vs 18 | Laparoscopic cholecystectomy in a pig model | BT versus VR trainer   | Performance score | GOALS five custom items | BT better than VR    |
| Denadai et al.   | 12 vs 12 | Simple interrupted sutures and subdermal interrupted sutures | Synthetic versus animal bench model | Performance score Efficiency | Global Rating Scale Mean number of stitches in given time | No significant difference |
| Hassan et al.    | 15 vs 15 | Laparoscopic psychomotor skill | BT versus VR            | Time              | Seconds            | VR better than BT   |
| Kanumuri et al.  | 8 vs 8   | Laparoscopic suturing and knot tying | Computer enhanced versus VR trainer | Time              | Seconds            | No significant difference |
| Kotari et al.    | 11 vs 13 | Laparoscopic psychomotor skill | Pelvic trainer versus VR trainer | Time              | Seconds            | No significant difference |
| Lehmann et al.   | 12 vs 12 | Laparoscopic psychomotor skill | Conventional video trainer versus VR trainer | Time              | Seconds            | No significant difference |
| Lin et al.       | 40 vs 20 | Laparoscopic suturing and knot tying | Direct vision training versus video-based training | Time              | Seconds            | Video-based better than BT |
| Madan and Frantziades | 14 vs 17 | Laparoscopic psychomotor skill | Conventional video trainer versus VR trainer | Performance score | Custom metrics Seconds | No significant difference |
| Nemani et al.    | 6 vs 7   | Laparoscopic pattern cutting task | Fundamentals of Laparoscopic Surgery (BT) versus Virtual Basic Laparoscopic Skills Trainer (VR) | Performance score | Custom metrics Seconds | No significant difference |
| Tanoue et al.    | 20 vs 20 | Laparoscopic psychomotor skill Suturing and knot tying | BT versus Procedicus MIST (VR) | Efficiency* Time | Errors Movements Seconds | BT better than VR |
| Van Bruwaene et al.| 21 vs 21 | Laparoscopic suturing | BT versus VR | Time | Seconds | BT better than VR |
| Youngblood et al.| 17 vs 16 | Three laparoscopic tasks in a pig model | Tower Trainer (BT) versus LapSim (VR) | Performance score | Custom global score Seconds | VR better than BT |

BT, box trainer; VR, virtual reality.

**Sources and search strategy**

We searched PubMed, Scopus and ERIC electronic databases. For PubMed, the search strategy used different combinations of Medical Subject Headings (MeSH) terms like (simulation training(MeSH Terms)), (surgery(MeSH Terms)), (education, medical, undergraduate(MeSH Terms)) with free text strings like “suturing”, “knot”, “basic skill”, “fidelity”. For Scopus and ERIC, we adopted a similar strategy based on the same keywords.

We also used the references of the retrieved reviews to look for other possible relevant articles.

The search was updated until June 2020. The result of the queries was uploaded into the Zotero reference management system.

**Study selection**

After removal of duplicates, the two authors (FC and GP) independently checked each title and abstract to identify eligible studies. We retrieved then a full-text copy of each selected article for a final decision on the fulfillment of the eligibility criteria. Any disagreements on the selection were resolved by discussion among the authors.

**Data collection process**

We analysed each selected article to extract the relevant information, according to the Patient Intervention Comparison Outcome model and the information was recorded in a form. In this meta-analysis, the subjects were undergraduate students; as already defined, the intervention was the description of the used device, the comparison was the level of resemblance, and the outcome was the set of trained skills, with their measure. When a study reported more than one outcome (eg, both score and time), the outcomes were separately recorded. When a study reported more than one measure for the same outcome, all the values were used to compute the overall effect size if they pertained to different tasks. If they were elements of a global score (like in GOALS score), only the total score was recorded. Finally, data were entered into the Review Manager software (RevMan V.5) by Cochrane collaboration.

**Risk of bias**

The quality of the evidence provided by the selected articles was assessed according to the GRADE guidelines, 29 embedded in the RevMan software. The authors independently assessed the articles using the GRADE rules (study limitations, imprecision, inconsistency, indirectness and publication bias) and computing the score according to the four GRADE quality levels, from high to very low. Any disagreement on the evaluation of quality of an article of one grade or more were resolved by discussion among the authors.

The publication bias was assessed with the funnel plot, observing if the distribution of the outcome measures was symmetrical around the midline.

**Synthesis of results**

We classified the studies according the two outcomes of scoring system and time for a task. The synthesis was computed as the standardised mean difference of effect. Variability was expressed as 95% CI. We adopted the random effect model because we expected a moderate to high heterogeneity among studies. Heterogeneity was expressed with the I² statistic, which describes the percentage of variation across studies. It has been suggested that a value below 40% represents low inconsistency; between 40% and 60% represents moderate inconsistency; and over 60% represents high inconsistency. We used the RevMan V.5 software for all statistical calculations and for the generation of the forest plots.
Four studies measured the skill as the outcome with a scoring system. Two studies reported the superiority of low resemblance; one study reported the superiority of low resemblance; the other two could not show any significant difference.

The pooled effect size (figure 2) was small and did not show any significant advantage of a simulation with a high level over a low level of resemblance (−0.19, 95% CI −0.44 to 0.06). Heterogeneity was moderate (I^2=42%).

Comparison of time for a task

Ten studies measured time to complete a task as the outcome. All of them expressed the measure in seconds for the task, but tasks were rather different: simple tasks like touching a target with a grasper or cutting a shape, or more complex like a part of a procedure (place a stitch with two ties). For this reason, the heterogeneity was high (I^2=79%). Three studies reported a significant larger effect size for high resemblance; one study reported the superiority of low resemblance; the other two could not show any significant difference.

The pooled effect size (figure 3) was small and did not show any significant advantage of a simulation with a high level over a low level of physical resemblance (−0.14, 95% CI −0.54 to 0.27).

RESULTS

Flow of search

Our initial search found 2091 articles. One hundred and one met the inclusion criteria after the first screening, but after reading the abstract and the full text, only 12 were eligible for the meta-analysis. They were all written in English. Figure 1 shows the flow of search and the reasons to discard the articles not fulfilling the inclusion criteria. Nine studies reported time as the outcome; three reported a scoring system; and two reported both outcomes.

Study characteristics

Overall, the selected 12 studies were randomised trials, reporting on 373 undergraduate students (mean of subjects 15.54±6.89). All the students were defined as ‘novice students’, without any previous experience of surgery. The four studies reported a scoring system as the outcome, measured from one to six different items; the 10 studies reported time for a task measured the time for one to four different tasks. Hence, the total number of measures included in the meta-analysis was 456 for score and 504 for time.

Five studies trained simple laparoscopic psychomotor skill, like pick and place jacks or thread the rings; five studies trained basic laparoscopic tasks (suturing and knot tying or cutting); one study trained part of a laparoscopic cholecystectomy; and one study trained simple interrupted and subdermal interrupted sutures. Table 1 summarises the characteristics of the included studies.

Study quality

Overall, the quality of the selected studies was from moderate to high: participants were randomised; we could not detect selection bias; some studies reported two different outcomes; other studies reported incomplete data for one of the outcomes, and this outcome was not considered in the meta-analysis. These studies were anyway considered for the outcome with complete data. A frequent bias was that the blinding of outcome assessment was not reported. The sample size was generally small; only two studies enrolled more than 40 students. The distribution of outcomes in the funnel plot was symmetric, both for scoring system and for time, so we could infer the absence of a publication bias.

Comparison of scoring systems

Four studies measured the skill as the outcome with a scoring system. Two of these studies used a validated score (GOALS and the Global Rating Scale); two used an original scoring system. One more study used a scoring system, but the article did not report the detailed figures, so it was not possible to use it for the calculation of the pooled effect size. One study reported a significant larger effect size for high resemblance; one study reported the superiority of low resemblance; the other two could not show any significant difference.

The pooled effect size (figure 2) was small and did not show any significant advantage of a simulation with a high level over a low level of resemblance (−0.19, 95% CI −0.44 to 0.06). Heterogeneity was moderate (I^2=42%).

DISCUSSION

Our result suggests that the use of a higher or lower level of physical resemblance in simulation does not have a different effect on learning of the simple tasks that constitute the basic surgical skills. We found the same effect both on skill, measured from one to six different items; the 10 studies reported time for a task measured the time for one to four different tasks. Hence, the total number of measures included in the meta-analysis was 456 for score and 504 for time.

Five studies trained simple laparoscopic psychomotor skill, like pick and place jacks or thread the rings; five studies trained basic laparoscopic tasks (suturing and knot tying or cutting); one study trained part of a laparoscopic cholecystectomy; and one study trained simple interrupted and subdermal interrupted sutures. Table 1 summarises the characteristics of the included studies.

Study quality

Overall, the quality of the selected studies was from moderate to high: participants were randomised; we could not detect selection bias; some studies reported two different outcomes; other studies reported incomplete data for one of the outcomes, and this outcome was not considered in the meta-analysis. These studies were anyway considered for the outcome with complete data. A frequent bias was that the blinding of outcome assessment was not reported. The sample size was generally small; only two studies enrolled more than 40 students. The distribution of outcomes in the funnel plot was symmetric, both for scoring system and for time, so we could infer the absence of a publication bias.

Comparison of scoring systems

Four studies measured the skill as the outcome with a scoring system. Two of these studies used a validated score (GOALS and the Global Rating Scale); two used an original scoring system. One more study used a scoring system, but the article did not report the detailed figures, so it was not possible to use it for the calculation of the pooled effect size. One study reported a significant larger effect size for high resemblance; one study reported the superiority of low resemblance; the other two could not show any significant difference.

The pooled effect size (figure 2) was small and did not show any significant advantage of a simulation with a high level over a low level of resemblance (−0.19, 95% CI −0.44 to 0.06). Heterogeneity was moderate (I^2=42%).

Comparison of time for a task

Ten studies measured time to complete a task as the outcome. All of them expressed the measure in seconds for the task, but tasks were rather different: simple tasks like touching a target with a grasper or cutting a shape, or more complex like a part of a procedure (place a stitch with two ties). For this reason, the heterogeneity was high (I^2=79%). Three studies reported a significant larger effect size for high resemblance; one study reported the superiority of low resemblance; the other two could not show any significant difference.

The pooled effect size (figure 3) was small and did not show any significant advantage of a simulation with a high level over a low level of physical resemblance (−0.14, 95% CI −0.54 to 0.27).
at the beginning of a technical training, can produce a cognitive overload that is detrimental for an effective learning. This is a possible explanation for the finding of the studies that reported a better outcome for a low level of resemblance. On the contrary, the role of a simulation based on a high level of resemblance is fundamental in transferring basic skills to a full procedure, as shown by many studies at the residency level.

Hence, we are not arguing that the simulation of surgical procedures with advanced and immersive technology should not be used, but that the devices and the context should always be aligned with the expected learning outcomes and with the level of expertise of the trainee.

The main limit of this meta-analysis is the rather small sample size of the considered studies. The pooled sample size was of 373 students; the calculations were done on a higher number of measures, and this is why we adopted a random effect model that should compensate the heterogeneity.

Despite these limits, we believe that this meta-analysis offers a relevant and reliable indication on the use of simulation in the training of basic surgical skill for undergraduate students.

CONCLUSION
Simulation with a low level of physical resemblance and the related low cost is a viable option for teaching and learning basic surgical skill at the undergraduate level. Further research should address the identification of the other elements that can contribute to a successful simulation, like peer feedback and self-regulated learning.

Twitter Gianmarco Panzeria @twi_gianmarco

Contributors FC designed the meta-analysis; GP did the search; both authors scanned the retrieved articles, selected and uploaded the eligible ones and discussed the results. FC drafted the manuscript; GP revised and approved the text.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The only generated data are the retrieved articles that are stored in the Zotero bibliographic management system. The database is available upon request. Please specify the preferred format.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD Fabrizio Consorti http://orcid.org/0000-0001-7096-4428

REFERENCES
1 Kneebone RL. Twelve tips on teaching basic surgical skills using simulation and multimedia. Med Teach. 1999;21:571–5.
2 Davis CR, Toll EC, Bates AS, et al. Surgical and procedural skills training at medical school – a national review. Int J Surg. 2014;12:877–82.
3 Cumming A, Ross M. The Tuning Project for Medicine—learning outcomes for undergraduate medical education in Europe. Med Teach. 2007;29:636–41.
4 Chai AL, Matsushima K, Sullivan ME, et al. Acute care surgery education in US medical schools: a systematic review of the current literature and report of a medical student experience. J Surg Educ. 2020;77:316–22.
5 Trible CG, Walters D. Preparing for the Lurch into a surgery clerkship. Heart Surg Forum. 2019;22:E172–9.
6 Adamas-Rappaport WJ, Waer AL, Teeple MK, et al. A comparison of unguided vs guided case-based instruction on the surgery clerkship. J Surg Educ. 2013;70:821–5.
7 Neylan CJ, Nelson EF, Dumon KR, et al. Medical school surgical boot camps: a systematic review. J Surg Educ. 2017;74:384–9.
8 Miller S, Sherrer E, Hasty B, et al. Introductory surgical skills course: technical training and preparation for the surgical environment. MedEdPORTAL. 2014;10:1775.
9 Woron AJ, Bentz CM, Glass PW, et al. The evolving role of simulation in teaching surgery in undergraduate medical education. Surg Clin North Am. 2019;95:739–50.
10 Willis RE, Van Sickle KR. Current status of simulation-based training in graduate medical education. Surg Clin North Am. 2015;95:767–79.
11 Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27:10–28.
12 Cheng A, Lockey A, Bhangi F, et al. The use of high-fidelity manikins for advanced life support training—A systematic review and meta-analysis. Resuscitation. 2015;93:142–9.
13 Beal MD, Kinnear J, Anderson CR, et al. The effectiveness of medical simulation in teaching medical students critical care medicine: a systematic review and meta-analysis. Simul Healthc. 2017;12:104–16.
14 Li MM, George J. A systematic review of low-cost laparoscopic simulators. Surg Endosc. 2017;31:38–48.
15 Hamstra SJ, Breugem BJ, Rafter R, et al. Reconsidering fidelity in simulation-based training. Acad Med. 2014;89:387–92.
16 Aleman M, Wynn GR, Arulampalam T. Virtual reality training in laparoscopic surgery: A systematic review & meta-analysis. Int J Surg. 2016;29:85–95.
17 Guedes HG, Câmara Costa Ferreira ZM, Ribeiro de Sousa Leão L, et al. Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis. Int J Surg. 2019;61:60–8.

CONSORT F, Panzera G. BMJ Simul Technol Enhanc Learn 2021;7:422–427. doi:10.1136/bmjstel-2020-000797
