

# Effect of just-in-time simulation training on provider performance and patient outcomes for clinical procedures: a systematic review

Matthew S Braga,<sup>1,2,3</sup> Michelle D Tyler,<sup>1,2,3</sup> Jared M Rhoads,<sup>1</sup> Michael P Cacchio,<sup>1,4</sup> Marc Auerbach,<sup>5</sup> Akira Nishisaki,<sup>6</sup> Robin J Larson<sup>1,3,7</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjstel-2015-000058>)

<sup>1</sup>The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire, USA

<sup>2</sup>The Children's Hospital at Dartmouth, Lebanon, New Hampshire, USA

<sup>3</sup>Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

<sup>4</sup>Tuck School of Business at Dartmouth, Hanover, New Hampshire, USA

<sup>5</sup>Department of Pediatric Emergency Medicine, Yale School of Medicine, Yale, Connecticut, USA

<sup>6</sup>Department of Anesthesia and Critical Care, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

<sup>7</sup>VA Outcomes Group, Department of Veterans Affairs Medical Center, White River Junction, Vermont, USA

## Correspondence to

Dr Matthew S Braga, Children's Hospital at Dartmouth, Pediatric Critical Care Medicine, One Medical Center Drive, Lebanon, NH 03756, USA; [matthew.s.braga@hitchcock.org](mailto:matthew.s.braga@hitchcock.org)

Received 24 June 2015

Accepted 16 September 2015

Published Online First

5 October 2015

## ABSTRACT

**Background** Providing simulation training directly before an actual clinical procedure—or 'just-in-time' (JiT)—is resource intensive, but could improve both provider performance and patient outcomes.

**Objectives** To assess the effects of JiT simulation training versus no JiT training on provider performance and patient complications following clinical procedures on patients.

**Study selection** We searched MEDLINE, Cochrane Library, CINAHL, PsycINFO, ERIC, ClinicalTrials.gov, simulation journals indexes and references of included studies during October 2014 for randomised trials, non-randomised trials and before-after studies comparing JiT simulation training versus no JiT training among providers performing clinical procedures. Findings were synthesised qualitatively.

**Findings** Of 1805 records screened, 8 studies comprising 3540 procedures and 1969 providers were eligible. 5 involved surgical procedures; the other 3 included paediatric endotracheal intubations, central venous catheter dressing changes, or infant lumbar puncture. Methodological quality was high. Of the 8 studies evaluating provider performance, 5 favoured JiT simulation training with 18–48% relative improvement on validated clinical performance scales, 16–20% relative reduction in surgical time and 12% absolute reduction in corrective prompts during central venous catheter dressing changes; 3 studies were equivocal with no improvement in intubation success, lumbar puncture success or urological surgery clinical performance scores. 3 studies evaluated patient complications; 1 favoured JiT simulation training with 45% relative reduction in central line-associated blood stream infections; 2 studies found no differences following intubation or laparoscopic nephrectomy.

**Conclusions** JiT simulation training improves provider performance, but currently available literature does not demonstrate a reduction in patient complications.

## BACKGROUND

Healthcare providers from various backgrounds are often expected to master multiple complex procedures. Historically, learners gained these technical skills as described by the 'see one, do one, teach one' mantra. While clinical procedures have become increasingly complex in recent decades, the training has remained unstandardised and even haphazard, resulting in growing concern about deficits in learners' technical competence and the potential relationship to adverse patient outcomes.<sup>1</sup>

Systematic reviews of simulation training for procedures suggest that it improves procedural performance on actual patients, increases provider confidence and has the potential to reduce patient complications.<sup>2–6</sup> Despite these findings, the most effective and efficient methods of providing simulation training remain unknown. Of particular interest is the issue of ideal timing of both initial and refresher simulation training. Currently, teacher and learner time constraints dictate that most simulation training is scheduled irrespective to the timing of the actual procedures.

'just-in-time' (JiT) simulation training—training that is done immediately prior to the actual procedure—has been proposed to improve the simulation training process.<sup>7–10</sup> Although the idea of 'warming up' or 'rehearsing' before a procedure has shown promise in the sports and music literature,<sup>11–13</sup> this technique has not been widely adopted in healthcare. While some studies suggest JiT simulation training may be superior to no JiT training for improving provider performance on clinical procedures, the impact on patient outcomes is less clear and the settings in which it might be of greatest value are not known.<sup>7–10 14–19</sup> Two recent systematic reviews found 'warm-up' or 'rehearsal' training led to improved technical performance, but one included mostly low-quality observational studies and just two types of complex procedures,<sup>14</sup> and the other included only surgical procedures.<sup>20</sup> To date, no comprehensive review has rigorously evaluated the impact of JiT simulation training for procedures on outcomes assessed on actual patients, and across a range of procedures, providers and clinical environments.

## METHODS

### Review protocol

We drafted a protocol outlining our planned approach to addressing our research question. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines for reporting our methods and findings.<sup>21</sup> Our protocol, which remained unchanged, is available on request.

### Study eligibility criteria

We included studies that met the following prespecified criteria: (1) the design was a randomised controlled trial, non-randomised controlled trial or before-after study; (2) participants were healthcare providers or trainees who perform clinical procedures; (3) the intervention was JiT simulation



CrossMark

**To cite:** Braga MS, Tyler MD, Rhoads JM, et al. *BMJ Stel* 2015;1:94–102.

training, defined as a technique to replace or amplify real patient experiences with guided experiences<sup>22</sup> which occurred within 24 h prior to the anticipated clinical procedure; (4) the control group received no JiT training, which could include simulation training as long as there was no explicit or thoughtful effort to time the training to actual performance of the procedure; and (5) the study included one or more of our prespecified outcomes (see online supplementary appendix S1).

### Outcome measures

We prespecified two categories of outcomes—provider performance and patient complications; all outcomes had to be assessed based on procedures performed on actual patients. We chose to exclude the studies measuring performance on simulators (T1—educational laboratory), as our goal was to assess the effects of JiT training on higher levels of the Kirkpatrick model, T2 (patient care practices) and T3 (improved patient and public health).<sup>23 24</sup> Provider performance outcomes included procedure success rate, defined as full execution of all intended procedural steps, scores on validated performance scales, duration of procedure, number of attempts to success and need for corrective prompts. Patient complication outcomes included rates of procedure-specific adverse events and rates of procedural delay. We prespecified clinical procedure success rate as the primary outcome.

### Search methods

We searched MEDLINE (1946–2014), the Cochrane Library (all databases, Issue 10 of 12 October 2014), CINAHL (1984–2014), PsycINFO (1958–2014) and ERIC (1966–2014) using strategies developed with the guidance of two professional librarians. The search themes were ‘Simulation’, ‘Just-in-time’ and ‘Clinical competence’. No limits or language restrictions were used for the MEDLINE, Cochrane Library, CINAHL or PsycINFO searches. Terms were modified as needed for specific databases. Each search was last updated in October 2014 (see online supplementary appendix S2 for full strategies). We searched ClinicalTrials.gov (1997–2014) for unpublished studies. We also reviewed the entire index (since volume 1) of the journals *Simulation in Healthcare* and *Clinical Simulation in Nursing*. Reference lists of eligible studies were manually reviewed for additional studies not found by other search strategies.

### Study selection, data collection and quality assessment

Two reviewers each screened half of the studies identified by the electronic database searches based on titles and abstracts, excluding only clearly irrelevant records. The remaining studies were distributed among the four members of the study team such that each record underwent independent full-text review by two reviewers. After isolating the eligible studies, two reviewers independently extracted data from each study using a standardised data collection form, which was developed by consensus of four members of the research team (see online supplementary appendix S3 for details) and assessed methodological quality using both the modified Newcastle-Ottawa Scale (mNOS) and the Medical Education Research Study Quality Instrument (MERSQI)<sup>22 25 26</sup> (see online supplementary appendix S4). While the NOS was originally developed for assessing non-randomised trials/observational studies, the scale we used is a ‘mNOS’ which has been used in other published systematic reviews.<sup>2 3 5 27</sup> Disagreements were resolved by

discussion between primary reviewers; a third investigator was consulted when necessary.

### Analysis

#### Data synthesis and heterogeneity

Owing to the varied nature of the reported results, we summarised each study’s findings for our two outcome categories qualitatively. If a study reported more than one measure of provider performance or patient complications, we reviewed the quantitative findings to determine the predominant finding for the outcome category: favours JiT, favours no JiT training or equivocal. When studies reported multiple outcome measures, we used clinical outcomes rather than knowledge measures to determine the predominant finding. If multiple components of a validated performance scale were reported, we used the overall scoring to establish the outcome category. We assigned the label ‘equivocal’ when the reported finding or findings demonstrated no statistically significant difference between study arms, or when the variation among a set of findings was too great to confidently determine a favoured arm. The quantitative findings that informed the qualitative summary judgements for each study are shown in [table 3](#).

#### Assessment of reporting biases

In order to informally evaluate for publication bias, we ordered the studies by sample size to assess whether the smaller studies had consistently larger effect sizes.

#### Subgroup analyses

We prespecified three subgroup analyses based on the following variables: (1) provider level of training (resident vs attending vs nurse), hypothesising that earlier learners would benefit more from JiT training; (2) procedure complexity (bedside vs surgical), hypothesising that more complex procedures would be more subject to degradation of skill over time, as well as have increased risk of complications, and thus benefit more from JiT simulation training; and (3) location of procedure (outpatient vs inpatient vs intensive care unit (ICU)), hypothesising that these locations would be proxies for severity of illness and that sicker patients would have more difficult procedures and higher risk of complications.

#### Sensitivity analyses

We evaluated the impact of methodological quality on our overall summary impressions in three ways. First, for each domain of the two tools, we excluded studies that were deemed high risk of bias. Second, we excluded studies that had total scores less than four on the mNOS or less than the median of the Medical Education Research Quality Instrument previously described in a systematic review.<sup>22</sup> Third, we subgrouped the studies according to whether they were randomised trials versus non-randomised study designs.

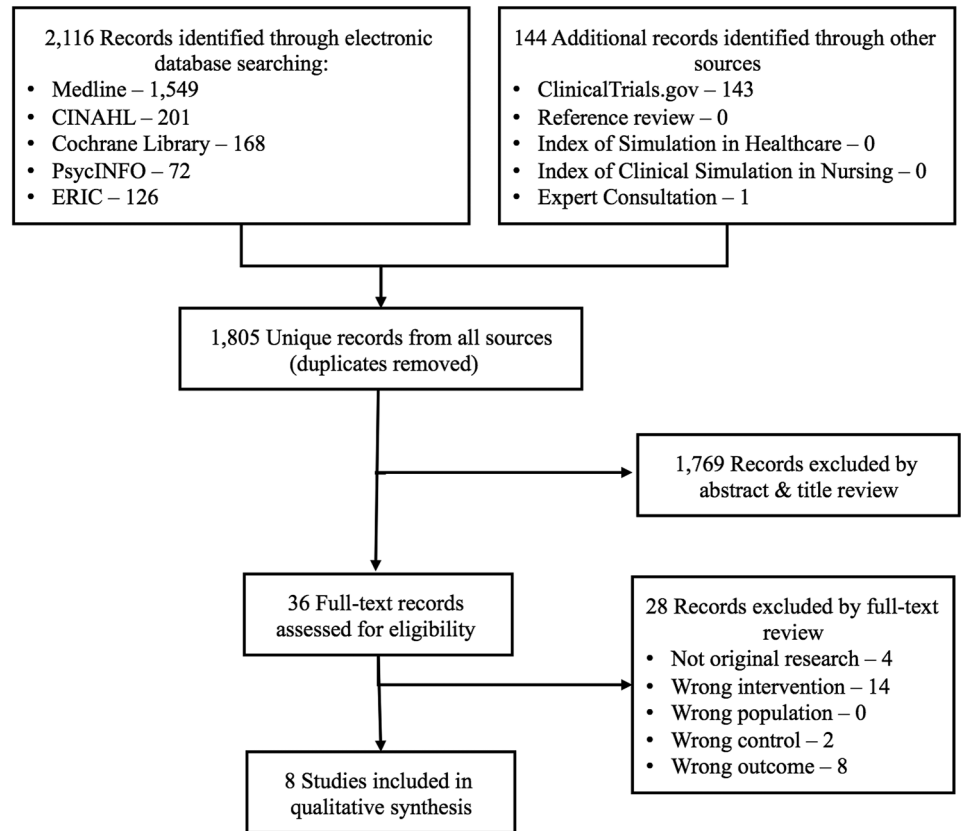
## RESULTS

### Description of studies

#### Results of search

As shown in [figure 1](#), our search strategies yielded 1805 potentially relevant publications. After exclusions based on title and abstract screening, 36 studies required full-text review, of which 8 met our inclusion criteria.<sup>7 9 10 15–19</sup> See [table 1](#) for study characteristics.

Figure 1 Study selection flow.



CINAHL = Cumulative Index to Nursing and Allied Health Literature  
ERIC = Education Resources Information Center

### Excluded studies

Two relevant studies were excluded from our review based on study design.<sup>28 29</sup> Willaert *et al*<sup>28</sup> found that patient-specific rehearsal prior to operation may be helpful in preoperative preparation and operative performance, and Desender *et al*<sup>29</sup> found that patient-specific rehearsal prior to endovascular aortic aneurysm repair to be feasible and to have face validity, but neither study design included a formal no JiT training control group.

### Methodological quality of included studies

The general methodological quality of the eight included studies was high based on both assessment tools. Just one study scored below 4 of 6 on the mNOS and none scored below 11 of 15 on the MERSQI (table 1).<sup>17 18</sup> Given the intervention being studied was training, blinding of the participants was not possible. Blinding of the assessor, however, was performed in five of the studies.<sup>7 9 15–17</sup> Two studies evaluated the intervention on a single surgeon, which limited the representativeness of the findings.<sup>17 18</sup>

### Overall provider performance

All eight studies reported at least one measure of provider performance, but each used unique measures for the specific procedure being studied. Measures included procedure success rate, components of validated performance scales including the Objective Structured Assessment of Technical Skill (OSATS)<sup>15</sup> and the Vassiliou scale (VAS),<sup>9</sup> surgical time, need for corrective prompts and number of attempts for success. Overall, five<sup>7 15 17–19</sup> of the eight studies favoured JiT simulation training

compared with no JiT training while three<sup>9 10 16</sup> studies were equivocal (table 2).

### Details of provider performance

#### Procedure success rate

Only two of the eight studies reported procedure success rates, both were non-operative bedside procedures.<sup>10 16</sup> Kessler *et al*,<sup>16</sup> who evaluated JiT versus no JiT training for paediatric and emergency room residents who perform infant lumbar punctures in the emergency room, reported no significant difference in overall success rates (38% vs 35%,  $p > 0.05$ ) but did report a decreased mean number of attempts ( $1.4 \pm 0.6$  vs  $2.1 \pm 1.6$ ,  $p < 0.01$ ). Nishisaki *et al*<sup>10</sup>, who evaluated JiT versus no JiT training for paediatric and emergency room residents who perform endotracheal intubations in paediatric ICU patients, reported no significant difference in intubation success rates based on first attempt (50% vs 62.5%,  $p = 0.44$ ) or overall (57.5% vs 75%,  $p = 0.19$ ; table 3).

#### Procedural time

Two studies used surgical time as a measure of procedural performance.<sup>9 18</sup> Lee *et al*<sup>9</sup> found no significant difference for the time to complete certain portions of various urological procedures including laparoscopic radical/partial nephrectomy, dismembered pyeloplasty and renal cyst decortication. Mucksavage *et al*<sup>18</sup> reported statistically significant improvements in surgical time comparing JiT versus no JiT training for both partial nephrectomies (227 vs 282 min,  $p = 0.004$ ) and radical nephrectomies (243 vs 290 min,  $p = 0.02$ ; table 3).

**Table 1** Characteristics of included studies

Study, year, country	Study design	Provider/learner	n	Procedure	n	Description of just-in-time intervention*	MERSQI (maximum 16)	mNOS (maximum 6)
Scholtz <i>et al.</i> <sup>19</sup> 2013, USA	Before-after study	Nurses	524	Central venous catheter dressing change (inpatient and outpatient)	2469	20 min of simulated central venous catheter dressing change with observation, assessment and debriefing, immediately prior to a dressing change	12.5	4
Nishisaki <i>et al.</i> <sup>10</sup> 2010, USA	Non-randomised control trial	Paediatric and emergency residents	78	Endotracheal intubation in PICU	401	10 min of hands-on orotracheal intubations on paediatric manikin and 20 min of multidisciplinary team training with high-fidelity infant simulator with debriefing, up to 24 h prior to intubation	12	4
Kessler <i>et al.</i> <sup>16</sup> 2015, USA	Before-after study	Paediatric and emergency residents	1319	Infant lumbar puncture	436	Standard Training PLUS 5–10 min of hands-on training on an infant lumbar puncture trainer coached by clinical supervisor, immediately prior to LP	14	4
Calatayud <i>et al.</i> <sup>15</sup> 2010, Denmark, UK, Canada	Randomised cross-over trial	Surgical residents	10	Laparoscopic cholecystectomy	20	15 min of virtual reality training of three tasks, immediately prior to surgery	14.5	6
Lee <i>et al.</i> <sup>9</sup> 2012, USA	Randomised cross-over trial	Urology residents	7	Laparoscopic radical and partial nephrectomy, dismembered pyeloplasty, renal cyst decortication	28	5 min of electrocautery simulation task training on virtual reality surgical simulator and 15 min of laparoscopic suturing on basic laparoscopic pelvic box trainer, approximately 1 h prior to surgery	14	5
Chen <i>et al.</i> <sup>7</sup> 2013, USA	Randomised control trial (randomised by surgical case)	Obstetric and gynaecological residents	29	Laparoscopic surgery for supracervical and total hysterectomies and adnexal/tubal surgery	91	Standard Training PLUS 15 min of low fidelity laparoscopic training of three tasks, immediately prior to surgery	15.5	6
Moldovanu <i>et al.</i> <sup>17</sup> 2011, Romania	Randomised control trial (randomised by surgical case)	Attending surgeon	1	Laparoscopic cholecystectomy	20	15–20 min of training on a virtual reality simulator with seven tasks, less than 15 min prior to surgery	11	5
Mucksavage <i>et al.</i> <sup>18</sup> 2012, USA	Before-after study	Attending surgeon	1	Laparoscopic partial and radical nephrectomies	75	15–20 min of training using pelvic suturing exercises, prior to surgery	11.5	3

\*All studies described control group as receiving 'no specific warm-up/training' except Chen *et al.*<sup>7</sup> (control=low fidelity training of 10 trials of three tasks at beginning of gynaecology rotation) and Kessler *et al.*<sup>16</sup> (control=viewed video and solitary training session on infant lumbar puncture trainer at start of residency until master performance standard achieved). MERSQI, Medical Education Research Study Quality Instrument; mNOS, modified Newcastle-Ottawa Scale; PICU, paediatric intensive care unit.

### Validated performance scales

Four studies, all evaluating operative/surgical procedures, used various validated scales to assess procedural performance.<sup>7 9 15 17</sup> Two of these studies, Calatayud *et al.*<sup>15</sup> evaluating JiT training among surgical residents performing laparoscopic cholecystectomy and Chen *et al.*<sup>7</sup> evaluating JiT training among obstetrics/gynaecology residents performing laparoscopic hysterectomy and adnexal/tubal ligation, used a modified OSATS to evaluate areas such as handling of equipment and tissue, use of assistance, and knowledge of procedure. Both studies found statistically significant differences in final OSATS scores favouring JiT over no JiT training (28.5 vs 19.3,  $p=0.042$  and 22.5 vs 19.5,  $p\leq 0.001$ , respectively, both out of 30 possible points<sup>7 15</sup>). Two studies used sections of the VAS score.<sup>7 9</sup> Chen *et al.*<sup>7</sup> used sections related to laparoscopic-specific dexterity, efficiency and tissue handling and found statistically significantly higher mean scores with JiT versus no JiT training (18 vs 15.5 out of 25 points,  $p\leq 0.001$ ). Lee *et al.*<sup>9</sup> evaluated JiT training among urology residents performing laparoscopic radical and simple pyeloplasty and cyst decortication; the study considered only the suturing and knot-tying portion of the scale and found no statistically significant difference between JiT and no JiT training

(3.50 vs 3.54, maximum score not provided,  $p=0.92$ ). Lastly, Moldovanu and colleagues evaluated JiT training for an attending physician performing laparoscopic cholecystectomy and combined elements of the OSATS and VAS to make a Global Rating Scale ranging from 6 to 30. They found improvement in the overall impression scores which ranged from 1 to 5, favouring JiT over no JiT training based on two evaluator assessments (4.86 vs 4.0,  $p=0.004$  and 4.86 vs 4.37,  $p=0.059$ )<sup>17</sup> (table 3).

### Other provider performance measures

Kessler *et al.*<sup>16</sup> also reported significant improvement in analgesia use prior to spinal needle insertion (68% vs 19%,  $p<0.001$ ) and a non-significant improvement in early (appropriate) stylet removal (69% vs 54%,  $p=0.06$ ). Scholtz *et al.*<sup>19</sup> evaluated JiT training for nurses who performed central line dressing changes and reported the percentage of dressing changes that required corrective prompts by observers was significantly lower following JiT versus no JiT training (9% vs 21%,  $p<0.001$ ; table 3).

### Patient complications

Three of the eight studies reported rates of procedural complications, one of which reported improvement after using JiT

**Table 2** Qualitative summary of key provider performance and patient complications for JiT versus no JiT training

Study (year)	Provider/learner	Procedure	Provider performance		Patient complications	
			Measure(s)	Favours	Measure(s)	Favours
Scholtz <i>et al</i> <sup>19</sup> (2013)	Nurses	Central venous catheter dressing change	Corrective prompts	JiT	Central line-associated blood stream infections	JiT
Nishisaki <i>et al</i> <sup>10</sup> (2010)	Paediatric and emergency residents	Endotracheal intubation in PICU	Success rate	Equivocal	Tracheal intubation-associated events	Equivocal
Kessler <i>et al</i> <sup>16</sup> (2015)	Paediatric and emergency residents	Infant lumbar puncture	Success rate	Equivocal	–	–
Calatayud <i>et al</i> <sup>15</sup> (2010)	Surgical residents	Laparoscopic cholecystectomy	OSATS <sub>7</sub>	JiT	–	–
Lee <i>et al</i> <sup>9</sup> (2012)	Urology residents	Laparoscopic radical and partial nephrectomy, dismembered pyeloplasty, renal cyst decortication	Vassiliou Scale	Equivocal	–	–
Chen <i>et al</i> <sup>7</sup> (2013)	Obstetric and gynaecological residents	Laparoscopic hysterectomy and adnexal/tubal surgery	OSATS <sub>6</sub> , Vassiliou Scale, Kundhal Scale	JiT	–	–
Moldovanu <i>et al</i> <sup>17</sup> (2011)	Attending surgeon	Laparoscopic cholecystectomy	OSATS <sub>3</sub> , Vassiliou Scale <sub>2</sub>	JiT	–	–
Mucksavage <i>et al</i> <sup>18</sup> (2012)	Attending surgeon	Laparoscopic partial and radical nephrectomies	Surgical time	JiT	EBL, Narcotic use, LOS, change in creatinine, ischaemic time	Equivocal

For each study, the preceding provider performance/patient complication measure was used to determine the predominant finding in the favours column (JiT, equivocal, no JiT). –, not reported; EBL, estimated blood loss; JiT, just-in-time; LOS, length-of-stay; OSATS, Objective Structured Assessment of Technical Skill (subscript indicates dimensions studied); PICU, paediatric intensive care unit; subscripts indicate dimensions studied.

simulation training and two of which found no difference (table 2).<sup>10 18 19</sup> Scholtz *et al*<sup>19</sup> reported significant improvement in central line-associated blood stream infections, which were reduced to 2.9/1000 line days after the implementation of JiT training compared with 5.3/1000 line days at baseline ( $p < 0.001$ ). On the other hand, Mucksavage *et al*<sup>18</sup> reported non-significant differences between patients whose attending surgeon performed JiT training prior to laparoscopic nephrectomy compared with patients whose attending surgeon did not perform the training for multiple variables including estimated blood loss, narcotic use postoperatively, length of hospital stay, change in creatinine levels and renal ischaemic time. Nishisaki *et al*<sup>10</sup> reported no significant difference in tracheal intubation associated events, 32.5% vs 20.8% (JiT vs no JiT, respectively,  $p = 0.4$ ).

#### Subgroup analyses

Subgrouping the studies by provider level found one study evaluated nurses (favoured JiT),<sup>19</sup> five studies evaluated residents (two favoured JiT, three equivocal)<sup>7 9 10 15 16</sup> and two studies evaluated attending (both favoured JiT).<sup>17 18</sup> Subgrouping by procedure complexity found three studies were done at the bedside (one favoured JiT, two equivocal)<sup>10 16 19</sup> and five studies were done in the operating room (four favoured JiT, one equivocal).<sup>7 9 15 17 18</sup> Subgrouping by location was not possible as location was often unreported. The number of contributing studies was low and variables differed across studies, making it difficult to isolate the effect of a single variable.

#### Sensitivity analyses

None of our sensitivity analyses based on methodological quality led to changes in our overall qualitative impressions, though the number of contributing studies was low.

#### Publication bias

After ordering the studies by size, no relationship between sample size and effect size was noted. Just as the varied outcome

measures used across different studies precluded quantitative pooling, it did not allow for construction of a funnel plot.

## DISCUSSION

### Summary of main results

This is the first review to describe the effect of JiT simulation training using a systematic approach, rigorous inclusion criteria and clinical outcomes. Although quantitative meta-analyses were not possible due to the heterogeneity of procedures and outcome measures, several important qualitative findings emerged. Overall, we found that compared with no JiT training, the use of JiT simulation training was an effective means of improving measures of procedural performance across a range of surgical and non-surgical procedures, performed by a variety of provider types. Unexpectedly, subgroup analyses revealed that for residents performing paediatric intubation and infant lumbar puncture—two low-frequency/high-acuity bedside procedures—improvement in success rates was not seen. Only three studies provided data for patient complications, among which just one found improvement with JiT simulation training. No direct harms of JiT simulation training, such as delayed procedure initiation, were mentioned or described in any of the studies.

Our goal was to evaluate the impact of JiT simulation training for procedures on outcomes assessed on actual patients and thus advance JiT simulation-based knowledge. In order to achieve this goal, we limited our inclusion criteria to only T2 and T3 level outcomes, knowing we would vastly decrease the number of eligible studies. We a priori accepted this anticipated decrease in quantity for an increase in the clinical relevance of the included studies. We believe it is useful to know that there are only eight studies, which evaluated the impact of JiT training on clinical outcomes.

### Overall completeness and applicability of evidence

Currently available high-quality studies provide a broad perspective on the impact of JiT simulation training on provider and

**Table 3** Details of provider performance and patient complication outcomes following JiT versus no JiT training

Study	Outcome measure	Effect size		p Value
		JiT	No JiT	
<i>Provider performance</i>				
<i>Procedure success rate</i>				
Kessler <i>et al</i> <sup>16</sup>	Successful lumbar puncture by trainee	38%	35%	>0.05
	Number of attempts, mean (SD)	1.4 (±0.6)	2.1 (±1.6)	<0.001*
Nishisaki <i>et al</i> <sup>10</sup>	Successful intubation by trainee—first attempt	50.0%	62.5%	0.44
	Successful intubation by trainee—overall	57.5%	75%	0.19
<i>Procedural time</i>				
Lee <i>et al</i> <sup>9</sup>	Time to complete mobilisation of colon (min)	26.6 (±12.0)	29.4 (±10.1)	0.4
	Time to complete intracorporeal suturing and knot tying to reapproximate the line of Toldt (min)	7.6 (±4.5)	5.8 (±2.9)	0.17
Mucksavage <i>et al</i> <sup>18</sup>	Laparoscopic radical nephrectomy, mean (SD)			
	Total operating room time (min)	334 (±66)	388 (±66)	0.010*
	Surgical time (min)	243 (±70)	290 (±61)	0.021*
	Laparoscopic partial nephrectomy, mean (SD)			
	Total operating room time (min)	320 (±75)	374 (±62)	0.050*
	Surgical time (min)	227 (±73)	282 (±60)	0.004*
<i>Validated performance scales</i>				
Calatayud <i>et al</i> <sup>15</sup>	OSATS Global Rating Scale†, median (range)	28.5 (18.5, 32.0)	19.3 (15.0, 31.5)	0.042*
Chen <i>et al</i> <sup>7</sup>	Reznick Scale/modified OSATS‡, mean (95%CI)	22.5§ (21, 24§)	19.5§ (18, 21§)	≤0.001*
	Vassiliou Scale¶, mean (95% CI)	18.0§ (19 to 20§)	15.5§ (14 to 17§)	≤0.001*
	Kundhal Scale**, mean (95% CI); lower score=superior performance	9.5§ (8.5 to 10.5§)	11.5§ (10.5 to 12.5§)	≤0.001*
Kessler <i>et al</i> <sup>16</sup>	Use of analgesia	68%	19%	<0.001*
	Appropriate stylet removal	69%	54%	0.06*
Lee <i>et al</i> <sup>9</sup>	Hand Movement Smoothness Score††, mean (variance)	0.73 (±0.05§)	0.46 (±0.05§)	<0.03*
	Tool Movement Score††, mean (variance)	0.73 (±0.03§)	0.58 (±0.05§)	<0.05*
	Posture Stability Score ††, mean (variance)	0.54 (±0.06§)	0.34 (±0.05§)	<0.05*
	Attention Score ††, mean (variance)	0.8 (±0.02§)	0.64 (±0.03§)	<0.02*
	Distraction Score ††, mean (variance)	0.34 (±0.01§)	0.53 (±0.05§)	<0.001*
	Mental Workload Score ††, mean (variance)	0.68 (±0.03§)	0.87 (±0.05§)	<0.02*
	Mobilisation of colon‡‡, mean (variance)	21.43 (±0.54)	19.86 (±0.51)	0.04*
	Intracorporeal suturing and knot tying to reapproximate the line of Toldt ‡‡, mean (variance)	3.50 (±0.23)	3.54 (±0.25)	0.92
Moldovanu <i>et al</i> <sup>17</sup>	Respect for tissue scores§§,¶¶—components of OSATS‡, mean (SEM)			
	Exposure of biliary region and adhesiolysis	4.71 (±0.18)	3.88 (±0.35)	0.059
		4.71 (±0.18)	3.88 (±0.29)	0.037*
	Dissection of cystic pedicle and critical view	4.43 (±0.29)	3.38 (±0.18)	0.008*
		4.71 (±0.18)	3.88 (±0.29)	0.037*
	Dissection of gall bladder	4.00 (±0.31)	3.00 (±0.18)	0.014*
		4.43 (±0.20)	3.63 (±0.32)	0.058
	Overall tissue score	4.43 (±0.20)	3.75 (±0.16)	0.021*
		4.57 (±0.20)	3.87 (±0.22)	0.041*
	Time and motion scores§§,¶¶—components of Vassiliou¶, mean (SEM)			
	Exposure of biliary region and adhesiolysis	4.29 (±0.18)	4.50 (±0.32)	0.593
		4.57 (±0.20)	4.13 (±0.35)	0.308
	Dissection of cystic pedicle and critical view	4.57 (±0.29)	4.00 (±0.26)	0.175
		4.71 (±0.18)	3.88 (±0.12)	0.002*
	Dissection of gall bladder	4.57 (±0.20)	4.38 (±0.26)	0.573
		4.57 (±0.20)	4.00 (0.19)	0.059
	Overall tissue score	4.71 (±0.18)	4.25 (±0.16)	0.081
		4.86 (±0.14)	4.00 (±0.19)	0.004*

Continued

Table 3 Continued

Study	Outcome measure	Effect size		p Value
		JiT	No JiT	
Overall impression scores§§,¶¶, mean (SEM)				
	Exposure of biliary region and adhesiolysis	<b>4.86</b> (±0.14)	4.50 (±0.27)	0.279
		<b>4.43</b> (±0.20)	4.25 (±0.25)	0.595
	Dissection of cystic pedicle and critical view	<b>4.71</b> (±0.18)	4.13 (±0.22)	0.069
		<b>4.71</b> (±0.18)	4.25 (±0.25)	0.169
	Dissection of gall bladder	<b>4.71</b> (±0.18)	4.25 (±0.16)	0.081
		<b>4.71</b> (±0.18)	4.13 (±0.22)	0.069
	Overall tissue score	<b>4.86</b> (±0.14)	4.37 (±0.18)	0.059
		<b>4.86</b> (±0.14)	4.00 (±0.19)	0.004*
Other measures				
Kessler <i>et al</i> <sup>16</sup>				
	Use of analgesia	<b>68%</b>	19%	<0.001*
	Appropriate stylet removal	<b>69%</b>	54%	0.06*
Scholtz <i>et al</i> <sup>19</sup>				
	Per cent of CVC dressing changes requiring corrective prompts to complete procedure	<b>9%</b>	21%	<0.001*
Patient complications				
Mucksavage <i>et al</i> <sup>18</sup>				
Laparoscopic radical nephrectomy				
	Estimated blood loss in mL, mean (SD)	<b>74.4</b> (±41.9)	130.9 (±186.4)	0.226
	Per cent of patients needing patient controlled analgesia	<b>12%</b>	28%	0.119
	Length-of-stay in hours admitted to hospital floor, mean (SD)	<b>65.6</b> (±31.1)	65.6 (±27.6)	0.995
	Change in Cr from baseline to discharge in mg/dL, mean (SD)	<b>0.46</b> (±0.34)	1.02 (±2.1)	0.287
Laparoscopic partial nephrectomy				
	Estimated blood loss in mL, mean (SD)	102.5 (±75.0)	<b>101.2</b> (±67.2)	0.960
	Per cent of patients needing patient controlled analgesia	20%	33%	0.662
	Length-of-stay in hours admitted to hospital floor, mean (SD)	<b>58.8</b> (±30.4)	66.4 (±39.4)	0.562
	Change in Cr from baseline to discharge in mg/dL, mean (SD)	0.13 (±0.22)	<b>0.12</b> (±0.14)	0.972
	Renal ischaemic time in minutes, mean (SD)	<b>29.4</b> (±9.76)	32.75 (±11.27)	0.533
Nishisaki <i>et al</i> <sup>10</sup>				
	TIAE—minor (% of cases)	<b>32.5%</b>	<b>20.8%</b>	0.4
	TIAE—major (% of cases)	0%	0%	—
Scholtz <i>et al</i> <sup>19</sup>				
	Central line-associated blood stream infection (per 1000 CVC line days)	<b>2.9</b>	5.3	<0.001*

Favoured arm is bold.

\*Statistically significant.

†OSATS is a validated Global Rating Scale based on seven components, scores range 7–35. Higher score=better performance.

‡Reznick Scale OSATS (modified) is a validated Global Rating Scale based on six components each scored on a five-point anchored Likert scale, scores range 6–30. Higher score=better performance. Also contains a pass/fail score.

§Estimated.

¶Vassiliou Scale is a validated Global Rating Scale based on five components each scored on a five-point anchored Likert scale, scores range 5–25. Higher score=better performance.

\*\*Kundhal Scale is a validated Global Rating Scale based on five components each scored on a five-point anchored Likert scale, scores range 5–25. Lower score=better performance.

††Scale ranges from 0 (no) to 1 (yes).

‡‡Modified Vassiliou Scale.

§§First line represents evaluator 1 scores, bottom line indicates evaluator 2 scores.

¶¶Global Rating Scale/Score.

Cr, creatinine; JiT, just-in-time; CVC, central venous catheter; OSATS, Objective Structured Assessment of Technical Skill; TIAE, tracheal intubation-associated events.

patient outcomes. However, the ability to draw firm conclusions for any one procedure, training approach, or provider group remains challenging. While the findings were generally consistent across the range of interventions and populations studied, we caution against generalising the findings to procedures and participants beyond those represented in our review. With regard to applicability, JiT training seems to be most appropriate as refresher training, with consideration for short training time and readily available relatively low-cost simulation equipment, we speculate that most interventions would be feasible; however, some of the interventions might be difficult to implement at scale due to the time and/or resources required. Although no delays in procedure initiation were described, we understand in high-acuity high-volume settings delays in care

could occur as a result of JiT training in a downstream ‘cascade effect’. While this ‘cascade effect’ has not been described as a consequence of JiT training, utilisers of JiT training must be cognisant of potential consequences the utilisation of time and resources may have on other patients.

### Quality of the evidence

The quality and rigour of the included studies were generally strong as indicated by MERSQI and mNOS scores. The directional effect of JiT simulation training on provider performance was relatively consistent; however, the varying nature of the procedures made it impossible to aggregate and summarise the findings in a concise, quantitative manner. The lack of assessor blinding in four of the included studies<sup>10 16 18 19</sup> could have

made JiT simulation training appear more favourable, as could have the single-surgeon intervention design in two of the studies.<sup>17 18</sup> It should be noted that secular trends were likely occurring during the Sholtz study, which may have contributed to the observed reductions in central line-associated blood stream infection rates.<sup>19</sup>

### Potential biases in the review process

We performed an extensive search of the literature and conducted all of our detailed screenings in duplicate fashion, with at least one experienced clinician in each pair. While using a single reviewer to perform the title and abstract screening could have resulted in missing one or more eligible studies, we sought to prevent this by only excluding obviously irrelevant studies at this step. The language used to describe the use of JiT simulation training is not standardised, which could have introduced bias by obscuring relevant studies. For example, other terms used to describe interventions similar or identical to JiT simulation training are 'refresher training', 'warm-up' and 'rehearsal'. JiT simulation training can also entail a variety of technologies, ranging from computer simulations and virtual reality trainers to physical manikins. Additionally, our reliance on qualitative analysis methods necessarily introduced some subjectivity, as it required us to make decisions about what effects were important. Finally, poor reporting of patient complications limited our ability to draw reliable conclusions.

### Agreements and disagreements with other studies or reviews

The fact that all of the eight included studies were published in the past 5 years suggests that JiT simulation training is a relatively novel concept, and the generally favourable findings likely forecast a trend towards increasing use over time. Additional factors likely to increase the use of JiT training include demand for improved efficiency and individuality in learner education, increased efforts to improve patient safety, and recognised need for ongoing refresher education.

The unexpected finding of no benefit for the two high-acuity/low-frequency procedure studies may be related to the fact that these studies focused on learners at the stage of establishing initial competency, while other successful studies were evaluating the impact of JiT training as refresher training or warm-up for procedures on which the providers had already established competency. In a recent article by Sawyer *et al.*,<sup>30</sup> an evidence-based pedagogical framework for procedural skill training is presented and discussed where learning is broken down into first a cognitive phase which includes 'learn' and 'see', followed by a psychomotor phase including 'practice', 'prove', 'do' and finally a maintenance phase. Our systematic review suggests that the current JiT training may be beneficial in the maintenance phase as a refresher but not adequate for the cognitive and psychomotor phases of initial procedural learning.<sup>30</sup> The included studies fit into the broader body of evidence supporting the idea that mentally visualising tasks and revisiting learned skills prior to performance can be beneficial.<sup>14</sup>

### CONCLUSIONS

We found that JiT simulation training leads to improved procedural performance in a variety of clinical situations, and while a hard link between these improvements and patient outcomes remains elusive, we believe that this evidence supports consideration of JiT simulation programmes and initiatives.

While continued and increased use of JiT training is justified based on procedural performance findings, we strongly advocate for further research including evaluation of patient-related outcomes. More research on the timing and duration of the training is needed to establish benchmarks for training programme design.

**Acknowledgements** The authors would like to recognise and thank the following individuals who contributed generously to this research effort: Kyla Donnelly, MPH; Thomas Mead, MLS; Heather Blunt, MLS.

**Contributors** MSB, MDT, JMR, MPC and RJL contributed significantly to all aspects of the study including conceptualisation and design, collection of the data, data analysis, drafting of the initial manuscript and reviewing of the final manuscript. MA and AN contributed significantly to the study including conceptualisation and design, data analysis and reviewing of the final manuscript.

**Competing interests** AN discloses funded research from the Agency for Healthcare Research and Quality (AHRQ). AN and MA are authors on manuscripts included in this systematic review.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** The authors are happy to share the data collected during manuscript review. They published most of the data collected.

### REFERENCES

- Singh H, Thomas EJ, Petersen LA, *et al.* Medical errors involving trainees: a study of closed malpractice claims from 5 insurers. *Arch Intern Med* 2007;167:2030–6.
- Madenci AL, Solis CV, de Moya MA. Central venous access by trainees: a systematic review and meta-analysis of the use of simulation to improve success rate on patients. *Simul Healthc* 2014;9:7–14.
- Kennedy CC, Cannon EK, Warner DO, *et al.* Advanced airway management simulation training in medical education: a systematic review and meta-analysis. *Crit Care Med* 2014;42:169–78.
- Cheng A, Lang TR, Starr SR, *et al.* Technology-enhanced simulation and pediatric education: a meta-analysis. *Pediatrics* 2014;133:e1313–23.
- Kennedy CC, Maldonado F, Cook DA. Simulation-based bronchoscopy training: systematic review and meta-analysis. *Chest* 2013;144:183–92.
- Ma IW, Brindle ME, Ronksley PE, *et al.* Use of simulation-based education to improve outcomes of central venous catheterization: a systematic review and meta-analysis. *Acad Med* 2011;86:1137–47.
- Chen CC, Green IC, Colbert-Getz JM, *et al.* Warm-up on a simulator improves residents' performance in laparoscopic surgery: a randomized trial. *Int Urogynecol J* 2013;24:1615–22.
- Kent DJ. Effects of a just-in-time educational intervention placed on wound dressing packages: a multicenter randomized controlled trial. *J Wound Ostomy Continence Nurs* 2010;37:609–14.
- Lee JY, Mucksavage P, Kerbl DC, *et al.* Laparoscopic warm-up exercises improve performance of senior-level trainees during laparoscopic renal surgery. *J Endourol* 2012;26:545–50.
- Nishisaki A, Donoghue AJ, Colborn S, *et al.* Effect of just-in-time simulation training on tracheal intubation procedure safety in the pediatric intensive care unit. *Anesthesiology* 2010;113:214–23.
- Fradkin AJ, Zazryn TR, Smoliga JM. Effects of warming-up on physical performance: a systematic review with meta-analysis. *J Strength Cond Res* 2010;24:140–8.
- Moorcroft L, Kenny DT. Singer and listener perception of vocal warm-up. *J Voice* 2013;27:258.e1–13.
- McHenry M, Johnson J, Foshea B. The effect of specific versus combined warm-up strategies on the voice. *J Voice* 2009;23:572–6.
- O'Leary JD, O'Sullivan O, Barach P, *et al.* Improving clinical performance using rehearsal or warm-up: an advanced literature review of randomized and observational studies. *Acad Med* 2014;89:1416–22.
- Calatayud D, Arora S, Aggarwal R, *et al.* Warm-up in a virtual reality environment improves performance in the operating room. *Ann Surg* 2010;251:1181–5.
- Kessler D, Pusic M, Chang TP, *et al.*, INSPIRE LP investigators. Impact of just-in-time and just-in-place simulation on intern success with infant lumbar puncture. *Pediatrics* 2015;135:e1237–46.
- Moldovanu R, Tărcoveanu E, Dimofte G, *et al.* Preoperative warm-up using a virtual reality simulator. *JSL* 2011;15:533–8.
- Mucksavage P, Lee J, Kerbl DC, *et al.* Preoperative warming up exercises improve laparoscopic operative times in an experienced laparoscopic surgeon. *J Endourol* 2012;26:765–8.
- Scholtz AK, Monachino AM, Nishisaki A, *et al.* Central venous catheter dress rehearsals: translating simulation training to patient care and outcomes. *Simul Healthc* 2013;8:341–9.



- 20 Abdalla G, Moran-Atkin E, Chen G, *et al.* The effect of warm-up on surgical performance: a systematic review. *Surg Endosc* 2015;29:1259–69.
- 21 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- 22 Cook DA, Hatala R, Brydges R, *et al.* Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011;306:978–88.
- 23 McGaghie WC, Draycott TJ, Dunn WF, *et al.* Evaluating the impact of simulation on translational patient outcomes. *Simul Healthc* 2011;6(Suppl):S42–7.
- 24 Smidt A, Balandin S, Sigafos J, *et al.* The Kirkpatrick model: a useful tool for evaluating training outcomes. *J Intellect Dev Disabil* 2009;34:266–74.
- 25 Reed DA, Cook DA, Beckman TJ, *et al.* Association between funding and quality of published medical education research. *JAMA* 2007;298:1002–9.
- 26 Wells GA, Shea B, O'Connell D, *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. Available from: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.htm](http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm) (cited 2014, October 22).
- 27 Cook DA, Brydges R, Zendejas B, *et al.* Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med* 2013;88:872–83.
- 28 Willaert WI, Aggarwal R, Van Herzele I, *et al.* Role of patient-specific virtual reality rehearsal in carotid artery stenting. *Br J Surg* 2012;99:1304–13.
- 29 Desender L, Rancic Z, Aggarwal R, *et al.* Patient-specific rehearsal prior to EVAR: a pilot study. *Eur J Vasc Endovasc Surg* 2013;45:639–47.
- 30 Sawyer T, White M, Zaveri P, *et al.* Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med* 2015;90:1025–33.