Core Competencies or a Competent Core? A Scoping Review and Realist Synthesis of Invasive Bedside Procedural Skills Training in Internal Medicine

Ryan Brydges, PhD, Lyna Stroud, MD, MEd, Brian M. Wong, MD, Eric S. Holmboe, MD, Kevin Imrie, MD, and Rose Hatala, MD, MSc

Abstract

Purpose
Invasive bedside procedures are core competencies for internal medicine, yet no formal training guidelines exist. The authors conducted a scoping review and realist synthesis to characterize current training for lumbar puncture, arthrocentesis, paracentesis, thoracentesis, and central venous catheterization. They aimed to collate how educators justify using specific interventions, establish which interventions have the best evidence, and offer directions for future research and training.

Method
The authors systematically searched Medline, Embase, the Cochrane Library, and ERIC through April 2015. Studies were screened in three phases; all reviews were performed independently and in duplicate. The authors extracted information on learner and patient demographics, study design and methodological quality, and details of training interventions and measured outcomes. A three-step realist synthesis was performed to synthesize findings on each study's context, mechanism, and outcome, and to identify a foundational training model.

Results
From an initial 6,671 studies, 149 studies were further reduced to 67 (45%) reporting sufficient information for realist synthesis. Analysis yielded four types of procedural skills training interventions. There was relative consistency across contexts and significant differences in mechanisms and outcomes across the four intervention types. The medical procedural service was identified as an adaptable foundational training model.

Conclusions
The observed heterogeneity in procedural skills training implies that programs are not consistently developing residents who are competent in core procedures. The findings suggest that researchers in education and quality improvement will need to collaborate to design training that develops a "competent core" of proceduralists using simulation and clinical rotations.

The American Board of Internal Medicine (ABIM)1 and the Royal College of Physicians and Surgeons of Canada (RCPSC)2 mandate that internal medicine (IM) physicians be competent in core invasive bedside procedures, yet neither provides formal guidelines for how to train learners to be procedurally competent. The absence of evidence-based standards for training and assessment is a critical gap because performing such procedures is not without risk. Procedural errors and complications can result in increased patient discomfort, longer hospital stays, and higher costs.3 Procedural complications are also a leading cause of adverse events identified in most national adverse event studies.4–9 A recent systematic review suggests that procedural complications result in 6.7% to 9.7% of hospital-wide adverse events and that nearly half of these events are considered preventable.10 IM programs require further guidance on the training model (or models) that develops internists who perform procedures competently.

With the competing demands of trainees needing opportunities to acquire skills and of patients expecting high-quality, safe health care, a delicate balance exists between medical education and patient safety. The majority of IM residents report performing fewer than five invasive bedside procedures during their undergraduate medical training,11 insufficient exposure to procedural skills during residency training,12 a lack of proficient faculty to supervise procedures,13 and low levels of comfort and confidence when performing procedures.14 These reports raise significant concern that IM programs currently offer inconsistent procedural skills training experiences that may lead to incompetent trainees and put patients at unnecessary risk. Supporting that concern, one survey showed that 70% of Canadian IM program directors agree that a national standard for assessing procedural competence would be beneficial.15 Improving procedural skills training can have broad implications for the health care system; for example, training interventions have been associated with reduced incidence of complications and preventable adverse events.16–18

To date, syntheses of the evidence on invasive bedside procedural skills training in IM include narrow systematic reviews on the use of simulation-based training for central venous catheterization19,20; a broad systematic review and meta-analysis of studies that are “heterogeneous and of varying quality and rigour”21; and a nonsystematic, critical synthesis presenting a broad conceptual framework for procedural skills training.22 We suggest there is a need for a comprehensive review using a systematic search paired with a technique
that is designed for synthesizing data from heterogeneous studies.

Accordingly, we conducted a scoping review and realist synthesis of this literature and aimed to characterize current procedural skills training interventions, collate how educators justify the interventions used in their programs, establish which interventions have the best evidence, and offer directions for future research and training. Our research question was: What can we learn from previous interventions designed to establish competence in five invasive bedside procedures (lumbar puncture, arthrocentesis, paracentesis, thoracentesis, and central venous catheterization [hereafter, the five invasive bedside procedures]) that are considered core competencies for internists in the United States and Canada?2

Method

We combined two complementary knowledge synthesis techniques: scoping review and realist synthesis. A scoping review is used to address "an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing existing knowledge."23 To provide an analysis beyond only describing the included studies, we used realist synthesis, which requires researchers to “unpack the context-mechanism-outcome relationship [associated with a training intervention], thereby explaining examples of success, failure, and various eventualities in between.”24

One explicit purpose of realist synthesis is to compare official expectations versus actual practice.24 In IM training, we suggest that the official expectation is that all residents develop competence in the five invasive bedside procedures. By contrast, we expected that our scoping review did not involve human participants, it was exempt from ethical review at all of the associated institutions.

Data sources, searches, and inclusion criteria

An experienced information specialist designed a peer-reviewed strategy (Supplemental Digital Appendix 1 at http://links.lww.com/ACADMED/A449)28 for us to use to systematically search the Medline, Embase, Cochrane Library, and ERIC (Education Resources Information Center) databases from inception to December 13, 2013. Search terms included medical subject headings and terms related to learner populations (e.g., medical student, resident, hospitalist, rheumatologist, fellows), to a training or assessment focus (e.g., certification, licensure, assessment), and to the procedures of interest (e.g., paracentesis, arthrocentesis). We updated our search on April 13, 2015, using the original terms, with searches in Medline and Embase only, as these two databases returned the greatest number and most specific results in our original search. We supplemented both searches by hand-searching the reference lists of published reviews and relevant journals, searching our own files, and consulting with colleagues who publish in the domain.

We included studies published in English language only. For participants, we included studies on any health care professional to capture procedural experience in multiple clinical settings (e.g., paracentesis in gastroenterology). We included studies across nonclinical (e.g., simulation laboratory) and clinical (e.g., academic hospitals) settings. We included training interventions focused on procedural skills training and/or on reorganizing clinical practices associated with invasive procedures. We chose the five invasive bedside procedures mentioned above because all are core competencies for internists in the United States and Canada;2 are performed frequently, and are linked to patient complications. We limited our focus to these five procedures to ensure synthesis feasibility. We included studies with any outcome of interest, such as measures of procedural competence, quality of care, and patient safety. With respect to study design, we included studies using any method (qualitative or quantitative), and both full-text articles and conference abstracts (collectively referred to as studies).

Study selection

We removed duplicate studies initially using Endnote software (Version X7.7.1, Thomson Reuters, New York, NY). We conducted study screening in three phases using DistillerSR (Web-based software, Evidence Partners, Ottawa, Ontario, Canada), performing all reviews independently and in duplicate. First, several of us (R.B., R.H., L.S., B.M.W.) pilot-reviewed 15 random study abstracts, reviewed consensus on the operational criteria for judging study inclusion, and discussed the adequacy of the search strategy. We then reviewed the abstracts of all studies and resolved conflicts by consensus (weighted kappa = 0.46, moderate agreement). Second, we reviewed all of the full-text articles, resolving conflicts by consensus. Interrater agreement (kappa) for the full-text review was 0.84 (original research reporting empirical data), 0.63 (focus on IM procedures), and 0.65 (focus on training). Third, we reviewed all of the studies, judging whether the context–mechanism–outcome linkage was a good representation or useful refutation of the proposed training intervention. This analysis resulted in our excluding many studies—for example, one in which the authors described the rationale as “a need to utilize novel technology” (context), designed a simulation-based intervention to improve physicians’ procedural competence (mechanism), and measured impact as teachers’ and learners’ perceived utility of training (outcome). We excluded this study on the basis of the superficial rationale, and the misalignment between the goal of training (i.e., improved trainee skills) and the performance outcome (i.e., self-report). In making these judgments, our interrater agreement was moderate (kappa = 0.47), and for the studies with disagreement between the two reviewers, we resolved conflicts via discussions with a third reviewer.

Quality assessment

Independently and in duplicate, we (R.B., R.H., L.S., B.M.W.) extracted information on learner and patient demographics, study design and methodological quality,
and details of the training interventions and measured outcomes. To evaluate study quality of the full-text articles (but not the conference abstracts), we used the MERSQI (Medical Education Research Study Quality Instrument). We resolved all conflicts via group consensus.

Data extraction
For studies with clearly described and well-aligned context–mechanism–outcome linkages, we used a structured form (Supplemental Digital Appendix 2 at http://links.lww.com/ACADMED/A449) to extract data for each component, which we operationalized as follows. Context included which cultural, institutional, societal, and/or regulatory drivers prompted the training intervention and details about the setting (or settings) in which the intervention was delivered. Mechanism included the theory or conceptual framework authors associated with their interventions and the process by which the intervention was expected to affect procedural skills competence. Outcome included authors’ proposed links between the outcomes they measured and the contexts and mechanisms they used, as well as the intended and unintended consequences associated with the intervention.

Data synthesis and analysis
After iterative team discussions, we (all authors) decided to aggregate studies according to mechanisms because organizing them this way better aligned with regulatory requirements from the RCPSC and ABIM, which do not attend to either context or outcome (i.e., they are usually focused on the training process), and resulted in a parsimonious set of four types of training interventions. We then performed our realist synthesis in three steps (outlined in Figure 1):

Step 1. In step 1, we synthesized findings for context, mechanism, and outcome separately for the studies categorized in each of the four intervention types. We followed the realist principle of seeking demiregularities (in this case, details that were consistently present or absent in authors’ descriptions) in how authors described their context, mechanism, and outcome.

Step 2. In step 2, we analyzed the results of step 1 and produced syntheses of the findings for context, mechanism, and outcome separately across all four intervention types.

Step 3. In step 3, we analyzed the collective syntheses from step 2, with the perspective that realist syntheses should “build explanations across interventions … that share similar underlying ‘theories of change’ as to why they work (or do not work) … in particular contexts.” Here, we aimed to identify a foundational training model that could be adapted to accommodate the lessons learned from all three steps of our realist synthesis.

Finally, we (all authors) presented preliminary results to a group of relevant stakeholders (26 IM program directors, residents, and researchers) during a conference in fall 2014 and discussed how the foundational training model we identified would interface with current training policies. After iterative rounds of consensus building, we produced a list of key components of future IM procedural skills training and developed a list of key lines of inquiry for future IM curriculum design and research.

Results
Study characteristics
Our initial and updated searches yielded 6,671 relevant studies, from which we identified 149 for full data extraction (91 full-text articles and 58 conference abstracts). Data on the characteristics of these 149 studies are included in Supplemental Digital Appendixes 3 and 4 (at http://links.lww.com/ACADMED/A449). Of these, we found that only 67 (45%) studies reported sufficient information about the context, mechanism, and outcome associated with the training intervention for realist synthesis. Figure 2 shows the flow of study inclusion or exclusion, and Table 1 provides a summary of study characteristics for all 67 included studies (63 full-text articles and 4 conference abstracts).

Realist synthesis findings
Our analysis yielded four types of procedural skills training interventions: “see one, do one,” educational-theory-informed (divided into mastery learning and other, including self-regulated learning and cognitive,
Theories70–77; medical procedural services (MPSs)78–86; and multifaceted quality improvement/patient safety (QI/PS) interventions.18,87–95 These four intervention types involved delivering procedural skills training in variable ways, and even within a single intervention type, studies described heterogeneous approaches to training. We describe each intervention type and the characteristics of the aggregated studies within each type in Chart 1.

Below, we outline the results of our syntheses from step 1 (within each intervention type) and step 2 (across all four intervention types) for context, mechanism, and outcome. We also provide a summary of our syntheses from step 3, which outlines our identified foundational training model.

Synthesis of context themes.
Step 1 (within each intervention type).
For all four intervention types, there was notable consistency across the contexts in which procedural skills training was initiated.

Authors used educational technologies, especially simulation, as the training modality in all intervention types, except for QI/PS interventions, which used in-service presentations and workshops grounded in clinical practice. Rationalies for using simulation included to capitalize on new educational technologies (see one, do one); to adhere to ABIM recommendations that simulation-based training should precede clinical practice (see one, do one); to move initial or early training away from patients, where harm may occur (see one, do one and educational-theory-informed); to evaluate the impact of educational designs, like competency-based education, on learning outcomes (educational-theory-informed); and to respond to the perceived decline of exposure to procedures during clinical training (see one, do one and MPS).

When simulation was used, authors mostly delivered training in simulation centers, with some “just in time, just in place” use of simulation in the clinical setting.

When procedural skills training took place in clinical settings, authors described a need to increase the quality of supervision from staff (MPS), as well as a need to avoid financial penalties—for example, from the Centers for Medicare and Medicaid Services, related to high infection rates (QI/PS).

Step 2 (across all four intervention types). Collectively, authors valued simulation as a safe training option for facilitating increased exposure to the repetitive practice of procedures. They positioned simulation as a precursor to, rather than a replacement of, clinical training and as a modality with which trainees can commit and learn from errors so that adverse events are minimized in clinical practice.

Synthesis of mechanism themes.
Step 1 (within each intervention type).
Despite the similarity in contexts, where authors’ rationales and study settings overlapped greatly, we found significant differences in how authors’ designed procedural training and in how they rationalized the underlying mechanism of training. For most see one, do one studies, authors suggested that novice trainees would experience increased comfort with and exposure to procedures via the active or experiential learning that technology-enhanced learning provides; yet, authors did not cite the
use of any educational principles in the design of training interventions. For educational-theory-informed studies, the mechanisms depended on authors’ chosen theory; for example, authors studying mastery learning proposed that baseline testing, deliberate practice with feedback, and a final assessment with a minimum passing standard would combine to ensure procedural competency. For most MPS studies, authors proposed that trainees would benefit from experiencing an integrated curriculum combining simulation-based training and clinical exposure on a two- or four-week procedural rotation. For most QI/PS studies, authors emphasized hospital-based rather than educational components, such as administrative and clinical champions (nurses were a common target group) who provided oversight, designed training interventions according to quality improvement principles (e.g., in-service presentations in the workplace), and served as drivers for accountability.

Step 2 (across all four intervention types). Collectively, all studies emphasized active learning during training, yet differed in how such active learning was accomplished: It was assumed in see one, do one studies, designed in educational-theory-informed studies, integrated in MPS studies, and situated in QI/PS studies.

Synthesis of outcome themes.

Step 1 (within each intervention type). For each of the four intervention types, there were significant differences in outcomes. For most interventions in nonclinical, simulation settings, authors assessed performance using individual-level outcomes such as self-reported confidence or direct observation of procedural skills. For most interventions in clinical settings, authors assessed performance using self-reported procedural success and group-level infection or complication rates, rather than direct observation of performance or chart audit. Often, authors did not report favorable validity evidence (e.g., reliability metrics) to demonstrate that patient outcomes were sensitive to the training interventions.

For see one, do one studies, the experimental groups’ procedural competence improved from baseline or

---

**Table 1**

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>No. of studies (n = 67)</th>
<th>No. of participants (n = 3,684)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-group, posttest only</td>
<td>2</td>
<td>289</td>
</tr>
<tr>
<td>Single-group pre/posttest</td>
<td>26</td>
<td>1,714</td>
</tr>
<tr>
<td>Nonrandomized, multiple groups</td>
<td>18</td>
<td>1,661</td>
</tr>
<tr>
<td>Qualitative</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>16</td>
<td>948</td>
</tr>
<tr>
<td>Conference abstract or could not tell</td>
<td>4</td>
<td>61</td>
</tr>
<tr>
<td><strong>Geographical location</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>55</td>
<td>3,339</td>
</tr>
<tr>
<td>Canada</td>
<td>8</td>
<td>201</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Asia</td>
<td>2</td>
<td>74</td>
</tr>
<tr>
<td>Oceania</td>
<td>1</td>
<td>Undefined</td>
</tr>
<tr>
<td>Conference abstract or could not tell</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td><strong>Study setting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonclinical setting (including simulation, computer-based, etc.)</td>
<td>37</td>
<td>1,582</td>
</tr>
<tr>
<td>Clinical setting (hospital, field, clinic)</td>
<td>8</td>
<td>219</td>
</tr>
<tr>
<td>Both</td>
<td>22</td>
<td>1,533</td>
</tr>
<tr>
<td><strong>Procedural skill studied</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumbar puncture</td>
<td>23</td>
<td>1,255</td>
</tr>
<tr>
<td>Arthrocentesis</td>
<td>4</td>
<td>196</td>
</tr>
<tr>
<td>Paracentesis</td>
<td>10</td>
<td>239</td>
</tr>
<tr>
<td>Thoracentesis</td>
<td>12</td>
<td>273</td>
</tr>
<tr>
<td>Central venous catheterization</td>
<td>43</td>
<td>2,889</td>
</tr>
<tr>
<td><strong>Ultrasound guidance used in training?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>30</td>
<td>1,294</td>
</tr>
<tr>
<td>No</td>
<td>36</td>
<td>2,355</td>
</tr>
<tr>
<td>Conference abstract or could not tell</td>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td><strong>Patient demographics reported?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>1,177</td>
</tr>
<tr>
<td>No</td>
<td>25</td>
<td>1,579</td>
</tr>
<tr>
<td>Not applicable</td>
<td>22</td>
<td>928</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical students</td>
<td>8</td>
<td>404</td>
</tr>
<tr>
<td>Physicians, postgraduate training</td>
<td>47</td>
<td>2,697</td>
</tr>
<tr>
<td>Physicians in practice</td>
<td>4</td>
<td>229</td>
</tr>
<tr>
<td>Nurses in practice</td>
<td>3</td>
<td>87</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>Mix of health care professionals, without a breakdown</td>
<td>13</td>
<td>44</td>
</tr>
<tr>
<td><strong>Clinical specialty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anesthesia</td>
<td>6</td>
<td>514</td>
</tr>
<tr>
<td>Critical care</td>
<td>14</td>
<td>1,230</td>
</tr>
<tr>
<td>Emergency medicine</td>
<td>16</td>
<td>950</td>
</tr>
<tr>
<td>Internal medicine, general or subspecialty</td>
<td>35</td>
<td>2,259</td>
</tr>
<tr>
<td>Family medicine</td>
<td>2</td>
<td>44</td>
</tr>
</tbody>
</table>

(Table continues)
as compared with control groups (who either had no training or traditional training). Most educational-theory-informed studies of mastery learning groups found that they outperformed control groups, though two large trials showed that a single mastery session did not improve future lumbar puncture success in pediatric patients. For other educational-theory-informed studies, authors applied most educational principles successfully (e.g., group conformity). All MPS groups improved from baseline or outperformed control groups, though authors commented that despite the observed benefits, the MPS was often assigned the most challenging patients, which may have implications for procedural success rates. For QI/PS studies, all showed improved outcomes related to the multifaceted approach, though none could specify which facet (or facets) led to the observed benefit, and none identified education as a key factor.

Step 2 (across all four intervention types). Although most authors labeled their training interventions as successful, our synthesis suggests that this was likely a function of their using weak comparator groups (e.g., nonintervention controls) and outcomes without sufficient evidence supporting their use as sensitive metrics. A clear demarcation was authors’ use of group-level assessments in clinical settings at the expense of individual-level assessments, like direct observation, which were used often in nonclinical, simulation settings.

Step 3: Collective syntheses summary and identified foundational training model. Together, our syntheses suggest a list of key components for future IM procedural skills training—namely, the need to design training that gives trainees the opportunity for active learning in a curriculum that tightly integrates simulation-based training with interventions in the clinical setting. The two intervention types that best aligned with these principles were the MPS and QI/PS. We suggest that studies of MPSs provide the most robust foundation for future procedural skills training curricula, particularly because most QI/PS studies did not describe the educational component of their multifaceted interventions adequately. Our analysis suggests that the MPS model will be adaptable to most institutional settings and can be customized to local settings using the lessons from our syntheses. For example, the MPS model can be adapted to address key context demirregularities by increasing both the volume and variability of training (e.g., training that varies relevant to situational or patient factors). Another adaptation is that the MPS model can be designed according to educational-theory-informed mechanisms to prompt active learning during simulation-based training. Finally, the MPS model can be adapted to include the notable practices identified in QI/PS studies, such as involving nursing and other health professions and identifying champions across clinical specialties.

We consider the implications of such adaptations to future MPS interventions and generate related key lines of inquiry below.

**Discussion**

We synthesized a heterogeneous literature to help stakeholders establish the key components of rigorous, evidence-based training for core invasive bedside procedures in IM. From 67 studies, we identified four intervention types, which we synthesized to identify key considerations for future IM procedural skills training curricula. The observed heterogeneity in how procedural skills training interventions are designed (mechanism) and in how competence is assessed (outcome) suggests that the official expectation that all residents develop competence in the five invasive bedside procedures is likely not fulfilled consistently. Our synthesis suggests that the most robust foundational model would be an adaptable MPS; this finding aligns with recent perspectives on procedural competence. After first

Table 1 (Continued)

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>No. of studies (n = 67)</th>
<th>No. of participants (n = 3,684)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab medicine, pathology, or radiology</td>
<td>2</td>
<td>430</td>
</tr>
<tr>
<td>Obstetrics-gynecology</td>
<td>1</td>
<td>242</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>17</td>
<td>701</td>
</tr>
<tr>
<td>Surgery</td>
<td>9</td>
<td>448</td>
</tr>
<tr>
<td>Only specified as a trainee class as captured above</td>
<td>4</td>
<td>164</td>
</tr>
<tr>
<td>Could not tell</td>
<td>2</td>
<td>74</td>
</tr>
</tbody>
</table>

**Outcome measures**

- Self-report, reaction, or confidence: 27, 1,492
- Knowledge: 23, 1,231
- Skill, time: 4, 138
- Skill, nontime: 25, 1,344
- Skill, product: 3, 122
- Behavior, time: 5, 222
- Behavior, nontime: 27, 2,263
- Patient outcome: 39, 2,629

**Study quality**

- All studies, mean = 12.7 (SD = 2.1): 62, 3,612
- MERSQI score > 12.7: 33, 2,680

Abbreviations: SD indicates standard deviation; MERSQI, Medical Education Research Study Quality Instrument.

*Note that of the 47 studies, 67 did not define the sample size of the learner population.
*Note that of the 13 studies, 12 did not define the sample size of the learner population.
*Note that skills are measured in the nonclinical setting, whereas behaviors are measured in the clinical setting. Nontime refers to skills and behaviors that are assessed based on observable processes (e.g., a Likert scale score assessing technical skill performance), while time refers to skills and behaviors that are assessed based on duration. For more information, see Chart 1.
*Skill, product refers to outcomes of performance in a nonclinical setting (e.g., successful task completion, or assessment of the quality of the product of task performance).
*Study quality was calculated using the MERSQI (total possible score of 18). To facilitate interpretation, note that a MERSQI score > 12.7

**Review**
### Chart 1
Select Characteristics of the Studies Aggregated Under Each of the Four Intervention Types in a 2015 Scoping Review and Realist Synthesis of the Literature on Invasive Bedside Procedural Skills Training in Internal Medicine

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“See one, do one” (n = 22)</td>
</tr>
<tr>
<td>Description</td>
<td>Interventions essentially replicated the see one, do one approach that is the hallmark of the clinical apprenticeship model.</td>
</tr>
<tr>
<td></td>
<td>Interventions were structured in a typical sequence: (1) lecture, (2) instructor demonstration, and (3) a short period of hands-on practice; very few ended with a performance-based assessment. Interventions used educational (Web-based [n = 4], simulation-based [n = 15]) or medical technologies (ultrasound guidance [n = 3]).</td>
</tr>
<tr>
<td></td>
<td>Most author groups did not cite educational evidence or theory as contributing to the instructional design.</td>
</tr>
<tr>
<td>Study design</td>
<td>Designs included two-group nonrandomized (n = 8), single-group pre/posttest (n = 7), RCT (n = 5), qualitative (n = 1), and conference abstract (n = 1).</td>
</tr>
<tr>
<td>MERSQI study quality score, mean (SD)*</td>
<td>12.7 (2.4), n = 20</td>
</tr>
<tr>
<td>Outcome measures**</td>
<td>Measures included trainee’s self-reported reactions to training or confidence (n = 13), tests of knowledge (n = 12), and patient outcomes (n = 11). Patient outcomes were measured via trainees’ self-reported procedural success (n = 4), observed performance or chart audit (n = 4), and group-level infection or complication rates (n = 3).</td>
</tr>
</tbody>
</table>

*Note that a MERSQI score could not be computed for five included full-text articles. To facilitate interpretation, note that a previous systematic review of 289 studies on simulation-based training found a mean MERSQI score (out of a possible 18) of 12.3 (SD = 1.8).

**The number of studies in this category may add up to more than the number of all studies because several studies reported multiple outcomes.

Abbreviations: MPS indicates medical procedure services; QI/PS, quality improvement/patient safety; RCT, randomized controlled trial; MERSQI, Medical Education Research Study Quality Instrument; SD, standard deviation.
Comparing our findings with those of previous reviews, we describe and consider the implications of three interrelated lines of inquiry for studying IM invasive bedside procedural skills training in the future.

Comparison with previous research

A recent systematic review on procedural skills training ended with recommendations to use simulation where possible; to use strong research designs like randomized controlled trials, especially when examining differences between instructional methods; and to teach using “competency-based methods such as mastery learning and deliberate practice.” While our review supports these suggestions, below we provide more specific recommendations for how simulation and clinical training can be integrated in a systems-based approach that aims to increase the volume and variability of procedural skills training opportunities and that emphasizes assessing evidence-based educational and clinical outcomes. Notably, our findings suggest that not all mastery learning interventions are successful, and thus future research will need to clarify the mechanisms for when mastery learning is successful or not, as well as the mechanisms of other promising instructional designs (e.g., directed self-regulated learning).

Line of inquiry 1

Based on the need to ensure accountability to patients and, by extension, regulatory bodies, as well as the limited training resources, and performance and observation of procedures in clinical practice, researchers need to test whether any adapted MPS model will be feasible for training all IM trainees.

Assuming the current level of resources and funding allocated to IM procedural skills training remains static, combined with IM residents’ limited clinical exposure to procedures, program directors will likely be challenged to implement any adapted MPS training model. If that assumption holds true, then policy makers may need to make the difficult decision to recommend targeted training of a smaller group of trainees, who have been identified as needing to develop and maintain procedural competence throughout their careers. A reinvestment of resources and training opportunities to smaller groups of trainees would mark a shift from expecting core competence in all trainees to training a competent core with a specialization in procedures. In such a system, for example, all IM residents could be expected to achieve cognitive competence (i.e., understand the indications, limitations, contraindications, and complications of procedures), as presently required by the ABIM. Beyond this cognitive competence, though, a proceduralist selection system would need to be implemented, based on trainee interest and a career path requiring procedural competence (e.g., plan to practice IM in community settings or in academic centers with a responsibility for training and assessment), to ensure a core set of clinicians who are procedurally competent. We acknowledge that this proposal would require large-scale changes in the procedure service-delivery models of hospitals that currently rely on all IM residents to perform procedures, as well as a philosophical shift in the professional identity and scope of practice of general internists.

A 2009 study provides a practical example of how programs might use criteria to decide privileges for performing procedures. When pulmonologists working at an outpatient pulmonary clinic learned that they had a higher frequency of iatrogenic pneumothorax compared with a nearby radiology practice, they imposed numerous practice changes including required retraining on thoracentesis skills to competency standards. The clinic did not allow pulmonologists who did not meet the standard to perform thoracentesis on patients. The authors reported a significant decline in pneumothorax rates, which held constant for two years post intervention. This example demonstrates the potential of investing in a core group of trainees, which could be a prudent resource allocation strategy that helps to address the pressing factors of system accountability, patient safety, and the rising costs of clinical errors. Research will be needed to determine whether this approach to training is appropriate for all invasive bedside procedures or whether trainee competence in some procedures might be realistically achieved in core training.

Line of inquiry 2

To build on and optimize implementation of adaptable MPS models, researchers will need to study how best to integrate the instructional designs of educational-theory-informed researchers and the systems-level thinking of quality improvement researchers.

While the MPS studies did use some notable practices of instructional design (i.e., integrating simulation with clinical training), they did not cite or use notable practices from QI/PS interventions (i.e., appointing champions and emphasizing accountability). A shortcoming of many QI/PS studies, however, was that they did not use simulation, which has been shown to be a common component of most procedural skills training interventions. Additionally, we found that authors of educational-theory-informed researchers should include systems-level QI/PS experts as team members in future studies, and hospital-based quality improvement teams should include education experts as members on their committees; both groups should work to align the design, implementation, and evaluation of procedural skills training that integrates the simulation and clinical settings.

Line of inquiry 3

Research is needed to evaluate validity evidence for outcomes measured in the nonclinical, simulation-based and clinical settings. Research generating evidence for relationships between patient and health care system outcomes and more accessible educational outcomes (i.e., educational surrogates) will be particularly important.

A 2013 article calls for research programs that establish evidence for links between outcome measures collected in the nonclinical setting with those collected in the clinical setting. For example, a 2015 meta-analysis examined the relationship between simulation-based assessments and clinical assessments.
and found that tools requiring raters to observe individual performance directly (e.g., global ratings of a procedure) showed the highest correlations between the two settings. 103 We suggest that the benefit of direct observation might result because assessing at the individual level helps avoid unit-of-analysis errors, which arise when outcomes are measured at a group level (i.e., collapsing infection or complication rates for an entire intensive care unit likely masks multiple data points from high or low performers, reducing the specificity of the measurement). Although there are approaches available to analyze such nested data, like hierarchical generalized linear models, 104 none of the studies included in our review adjusted for such nesting using these techniques.

Researchers will need to collect a wide array of validity evidence to clarify “pathways that link training interventions to patient health outcomes.” 105 Rather than using outcomes that are low-hanging fruit and for which there is little validity evidence, such as self-reported procedural success and group infection or complication rates, researchers will need to identify educationally sensitive outcomes in the clinical setting, especially those involving direct observation, 106, 107 and establish chains of evidence between outcomes measured in the nonclinical and clinical settings. 102, 103 Given the validity evidence supporting the use of global rating scales (with or without checklists) in the simulation setting, 108–110 adapting these scales to the clinical setting is likely a fruitful research direction.

Limitations

The primary literature on IM bedside procedural skills training had several limitations which impacted our review. Authors reported nearly universal success and few failures of their training interventions, which implies that there may be an issue of publication bias of positive studies in our dataset. Some procedures were studied more extensively than others, and nearly all studies emphasized the procedures’ technical components and excluded components such as judging whether a procedure needs to be performed, obtaining informed consent, coordinating care, and documenting the procedure. 111 All but one study 90 evaluated how training affects the development of procedural competence rather than the maintenance of competence. Although we judged the context–mechanism–outcome linkage independently and in duplicate, our evaluations remain subjective; however, that only 45% of studies met our standard for sufficient information on the context, mechanism, and outcome suggests that there are important gaps in how research on procedural skills training has been conducted and reported. By using a realist synthesis approach, we excluded many studies, some of which might have unearthed additional themes. We did not conduct meta-analyses, particularly because we believe that knowledge synthesis methods supported by qualitative research paradigms, like realist synthesis, provide more targeted answers regarding gaps in research, as well as potential solutions and next steps.

Conclusion

We found that actual practices in procedural skills training in IM are highly variable. Such variability is not surprising considering that regulatory organizations mandate procedural competence, yet do not provide guidelines for program directors to follow when implementing training programs. We have identified the MPS as a foundational training model and provided a list of potential key components that educators can incorporate into future procedural training curricula, which researchers can study and test systematically. In an era where evidence shows that high-quality training translates into high-quality care, 101, 104, 112 the imperative to design the best educational experience for our trainees has never been stronger.

Acknowledgments: The authors wish to thank Arija Birze and Judy Tran for providing research support, and Laure Perrier and Marina Englesakis for developing and conducting the literature search strategies.

Funding/Support: The authors are thankful for study funding provided by the Canadian Institutes of Health Research via Knowledge Synthesis Grant no. 201304KRS-132055.

Other disclosures: None reported.

Ethical approval: As this study did not involve human subjects, it was exempt from ethical review at all associated institutions.

R. Brydges is assistant professor, Department of Medicine, University of Toronto, and scientist, Wilson Centre, University Health Network, Toronto, Ontario, Canada.


101 Wong BM, Holmboe ES. Transforming the academic faculty perspective in graduate medical education to better align educational and clinical outcomes. Acad Med. 2016;91:473–479.


107 Carracco C, Englundar R, Holmboe ES, Kogin JR. Driving care quality: Aligning


