

# Screening residents for infant lumbar puncture readiness with just-in-time simulation-based assessments

David O Kessler,<sup>1,2</sup> Todd P Chang,<sup>3,4</sup> Marc Auerbach,<sup>5</sup> Daniel M Fein,<sup>6,7</sup> Megan E Lavoie,<sup>8,9</sup> Jennifer Trainor,<sup>10,11</sup> Moon O Lee,<sup>12</sup> James M Gerard,<sup>13,14</sup> Devin Grossman,<sup>15,16</sup> Travis Whitfill,<sup>17</sup> Martin Pusic<sup>18</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjstel-2016-000130>).

For numbered affiliations see end of article.

## Correspondence to

David O Kessler; Department of Pediatrics, Columbia University College of Physicians and Surgeons, 3959 Broadway, CHN-1-116, NYC 10032, USA; [drkessler@gmail.com](mailto:drkessler@gmail.com)

Accepted 13 October 2016  
Published Online First  
28 October 2016

## ABSTRACT

**Background** Determining when to entrust trainees to perform procedures is fundamental to patient safety and competency development.

**Objective** To determine whether simulation-based readiness assessments of first year residents immediately prior to their first supervised infant lumbar punctures (LPs) are associated with success.

**Methods** This prospective cohort study enrolled paediatric and other first year residents who perform LPs at 35 academic hospitals from 2012 to 2014. Within a standardised LP curriculum, a validated 4-point readiness assessment of first year residents was required immediately prior to their first supervised LP. A score  $\geq 3$  was required for residents to perform the LP. The proportion of successful LPs (<1000 red blood cells on first attempt) was determined. Process measures included success on any attempt, number of attempts, analgesia usage and use of the early stylet removal technique.

**Results** We analysed 726 LPs reported from 1722 residents (42%). Of the 432 who underwent readiness assessments, 174 (40%, 95% CI 36% to 45%) successfully performed their first LP. Those who were not assessed succeeded in 103/294 (35%, 95% CI 30% to 41%) LPs. Assessed participants reported more frequent direct attending supervision of the LP (diff 16%; 95% CI 8% to 22%), greater use of topical analgesia (diff 6%; 95% CI 1% to 12%) and greater use of the early stylet removal technique (diff 11%; 95% CI 4% to 19%) but no difference in number of attempts or overall procedural success.

**Conclusions** Simulation-based readiness assessments performed in a point-of-care fashion were associated with several desirable behaviours but were not associated with greater clinical success with LP.

## INTRODUCTION

Resident infant lumbar puncture (LP) success rates are poor (~35%).<sup>1–3</sup> Clear risks arise when under-prepared residents are allowed to perform LPs as LP failures can prolong hospital stays, antibiotics and/or antiviral use.<sup>4 5</sup> With repeated attempts, infants are also exposed to pain, bleeding, infection and prolonged periods of being restrained.

Competency in LP is expected from paediatricians. Trainees must be supervised until they can be entrusted to perform independently; however, clear standards to assess readiness for LP performance are lacking.<sup>6–8</sup> Supervisors often rely on a trainee's

self-described knowledge of the procedure to determine whether or not to allow their performance on a patient.<sup>8–10</sup> LP simulation can provide supervisors an opportunity to directly assess skills, though its role in assessing competency and readiness for supervised or independent performance at the point of care is largely unknown.<sup>11</sup>

The Lumbar Puncture workgroup of the International Network for Simulation-based Pediatric Innovation, Research, and Education (INSPIRE) has previously reported validity evidence for using a global rating scale to assess a trainee's readiness for LP.<sup>12 13</sup> Furthermore, we demonstrated an association between just-in-time simulation-based assessments in the workplace and clinical LP success.<sup>14</sup> In the prior study, these assessments were not used to decide who would or would not perform an LP. Therefore, we now prospectively test the just-in-time use of these assessments for aiding supervisors in determining whether or not to allow first year residents to perform their first supervised LP.<sup>14 15</sup> Our primary aim was to determine whether supervisor screening of residents using simulation-based assessments immediately prior to the first supervised attempt at infant LP would increase resident success. We hypothesised that screening residents for readiness would increase their successful first needle insertion beyond the rate of 38% (95% CI 34% to 43%) established by our prior cohort of 399 first year residents.<sup>16</sup> We anticipated that the assessment could increase LP success rates in one of two ways: either low-performing participants would not be entrusted to perform the LP, or residents would practise on a simulator with their supervisor becoming better prepared for the imminent procedure. Our secondary hypotheses were that we would see improved overall success with the procedure (on any attempt), a decrease in the number of attempts, greater use of analgesia and the early stylet removal technique (advancing the needle with stylet removed after puncturing the skin).<sup>17</sup>

## METHODS

### Study design and setting

This multi-institutional, prospective, phased educational cohort study was conducted over two consecutive academic years (2012–2014) in 35 academic hospitals within INSPIRE (see online supplementary content: table S1). All 35 sites obtained approvals from institutional review boards and, when required, informed consents were obtained.



CrossMark

**To cite:** Kessler DO, Chang TP, Auerbach M, et al. *BMJ Stel* 2017;**3**:17–22.

**Figure 1** Global Skills Assessment (GSA) Scale.

Novice (1)	Beginner (2)	Competent (3)	Proficient (4)
Extensive guidance More than 2 prompts	Minimal guidance 1 or 2 prompts	Performed independently Self corrected, made minor errors, or seemed hesitant	Performed independently No prompts, no errors

Prompt = verbal interjection required to either prevent or correct an error.

## Participants

Two consecutive cohorts of postgraduate year 1 residents were eligible for study participation during their first academic year. This included paediatric, emergency medicine and other first year residents (eg, combined track) depending on the site.

## Intervention

### Standardised procedure curriculum

At orientation prior to first year of residency, all residents completed a standardised published LP curriculum composed of watching videos, demonstrating performance on a simulator (BabyStap Laerdal Medical, Stavenger, Norway) and having individually coached practice sessions until a predefined mastery performance standard was achieved.<sup>16 18 19</sup>

### Just-in-time readiness assessments

We intended that a just-in-time readiness assessment would occur immediately prior to each resident's first LP opportunity in the clinical setting. The resident was to perform a mock LP on a simulator while the supervisor observed and completed a global rating scale based on the resident's performance (figure 1). This abbreviated scale shows strong correlation with a longer subcomponent checklist of critical skills and there is evidence to support its discriminant validity and high inter-rater reliability between attending raters.<sup>13</sup>

On the basis of our prior data, we proposed to the supervisors that a minimum score of 3 on the four-point scale was the prerequisite to allow residents to perform under supervision.<sup>14</sup> Residents and their clinical supervisors were encouraged to have a practice refresher on the simulator prior to assessment and to repeat the assessment as needed. Residents who did not achieve a passing score on the simulation assessment were instructed not

to perform the clinical procedure. Residency programme directors at 21 sites signed formal attestations in support of this requirement.

## Faculty development

Clinicians who supervise resident LPs were sent an online training module describing how to conduct readiness assessments. Module access was tracked for academic year 2013–2014 using unique logins on a learning management system (Moodle, Perth, Western Australia, Australia). Supervisors were senior residents, fellows, attending physicians and mid-level providers depending on the institution. We defined 'trained' supervisors as those completing modules, 'untrained' as documented lack of completion and 'unknown' for those with no documentation.

## Comparisons

We defined two groups for comparison to our group that experienced the intervention as intended: (1) a historical group of residents from the 2 years prior (2010–2012) who with similar LP training but no restriction from clinical performance by a threshold passing score; (2) the group of residents from the same 2 years (2012–2014) who performed a clinical LP but did not participate in the intended readiness assessment (non-assessed participants).

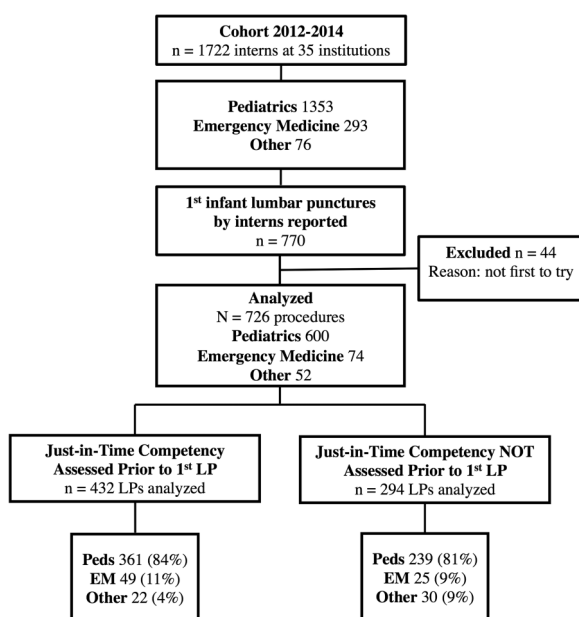
## Main outcome(s) and measure(s)

The primary outcome was procedural success, defined as obtaining cerebrospinal fluid (CSF) on the first needle insertion with <1000 red blood cells/high powered field or, if no cell count was available, clear CSF.

Other process measures were success on any attempt (as opposed to first attempt), the total number of attempts and the use of analgesia and early stylet removal technique (advancing the needle with stylet removed after puncturing the skin). We also noted family member presence and presence of attending physician as supervisor during the clinical procedure. Exploratory analysis was performed to look for any additional factors associated with LP success. Data were collected using an online self-report data form. Residents were asked to report data from their first LP experience, regardless of whether or not they participated in the readiness assessment. Site directors sent monthly reminder emails to residents with a link to the form. We evaluated implementation by measuring adherence to the protocol over both years of the study, specifically (1) whether or not the training supervisors had completed the supervisor training module and (2) whether or not the assessment was carried out.

## Analysis

Participant characteristics were summarised using standard descriptive statistics. Comparison of the per protocol study cohort to the historical group and to those who did not complete readiness assessments was performed using the  $\chi^2$  test. Process measures and implementation were analysed using  $\chi^2$  tests or Z-tests for proportions, and independent two-sided t-tests or Mann-Whitney U test for continuous variables, where appropriate. All analyses were conducted using SPSS V.22.0 (IBM, Armonk, New York, USA).



**Figure 2** Study flow diagram. EM, emergency medicine; LP, lumbar puncture; Peds, paediatrics.

To explore which variables were associated with success, we used a multivariable, mixed-effects, binary logistic regression model to determine the odds of achieving a successful LP controlling for variables that differed significantly between the residents who participated in the assessment versus those who did not (ie, residency type, location of LP, supervisor and patient age). The mixed-effects model also adjusts for intersite variability that could potentially confound the relationship between the variables and LP success.

## RESULTS

A total of 770 LPs were reported from 1722 residents (45%) with 44 excluded due to incomplete data, leaving 726 LPs for analysis (figure 2).

### Main outcome measure—LP success

Pooling all residents (2012–2014), success was achieved on the first attempt in 277 of 726 cases (38%, 95% CI 35% to 42%). Those residents who participated in the readiness assessment as intended (per protocol) were successful on the first attempt in 174/432 LPs (40%, 95% CI 36% to 45%). This was similar for both years of the study with 95/248 (38%, 95% CI 32% to 44%) successful in the first year of enrolment and 79/184 (43%, 95% CI 36% to 50%) successful in the second year of the study. The 40% success rate among participants in the assessment was similar to that of our historical group where 152/399 (38%) were successful on the first attempt (95% CI diff -4% to +9%).<sup>16</sup> In our study cohort, there was no statistical difference in success between those who participated in the readiness assessment (per protocol) and those who did not participate in the readiness assessment as intended (40% vs 35%; 95% CI diff -2% to +12%).

Table 1 compares factors between assessed and non-assessed participants. Procedures reported by assessed participants were more likely to: occur in the emergency department setting, have technicians as the holder for the procedure and be performed on older infants (table 1).

In our multivariable, mixed-effects, binary logistic regression model, no clear modifying factors of success were identified between those who did and did not complete an assessment after an adjusted analysis (table 2).

### Other process measures

With the less conservative definition of success (on any attempt, not just the first one), residents were successful 330 out of 726 times (45%, 95% CI 42% to 49%) with a median of 1 attempt (IQR 1–2). Table 3 compares procedural factors between assessed and non-assessed participants, with the former having greater use of the early stylet removal technique and topical analgesia. Non-assessed participants were also more likely to have had direct attending supervision.

### Pathway implementation

Supervised readiness assessments were reported in 432 of the 726 LPs (60%). There was improvement in pathway adherence between the first and second years of this new intervention with 248/441 (56%) adhering in 2012–2013 and 184/285 (65%) adhering in 2013–2014 (diff. 8%; 95% CI 1% to 15%). The supervisors for procedures performed by assessed participants were more likely to have undergone the supervisor training; however, most LPs were supervised by faculty who had not undergone the specific training for the readiness assessment.

## DISCUSSION

Employing just-in-time readiness assessments (our intervention) did not lead to an improvement on the first attempt or overall LP success rates among residents when compared with both a historical group and to the non-assessed participants (residents who did not complete the readiness assessment prior to performing an LP). However, compared with the latter, assessed participants had greater attending clinical supervision and displayed higher rates of adaptive behaviours such as the early stylet removal technique and topical analgesia use.<sup>17</sup> The assessment process was feasible across a large number of institutions and participation in the readiness assessments improved over the 2 years of the study (56% → 65%).

Even experienced providers (>100 LPs) only achieve atraumatic success with infant LP in ~70% of cases.<sup>5</sup> For residents, we have previously reported on their generally low prior experience and poor performance of LP in clinical settings.<sup>2 3 14 16 20</sup> The success rate of 43% seen in the final year of this study is slightly higher than those achieved in any of our prior studies (34–42%); however, this is still a disturbingly low success rate.

**Table 1** Factors associated with participation in readiness assessments (2012–2014)

Factors	Readiness assessed* (n=432)	Readiness not assessed* (n=294)	p Value†
<i>Provider variables</i>			
<i>Specialty</i>			
Paediatrics	361 (83.6)	239 (81.3)	NS
Emergency medicine	49 (11.3)	25 (8.5)	NS
Medicine/paediatrics	19 (4.4)	25 (8.5)	0.02
Other	3 (0.7)	5 (1.7)	NS
Overall	–	–	0.02
<i>Location where LP performed</i>			
ED	247 (57.3)	87 (29.6)	<0.001
PICU	8 (1.9)	5 (1.7)	NS
NICU	65 (15.0)	99 (33.7)	<0.001
Inpatient	100 (23.1)	89 (30.3)	0.03
Other	12 (2.7)	14 (4.7)	NS
Overall	–	–	<0.001
<i>Holder for patient</i>			
Attending	17 (3.9)	9 (3.1)	NS
Nurse	214 (49.5)	181 (61.6)	0.001
Resident	39 (9.0)	25 (8.5)	NS
Fellow	9 (2.1)	14 (4.8)	0.04
Tech/nursing assistant	136 (31.5)	53 (18.0)	<0.001
Others	17 (3.9)	12 (4.1)	NS
Overall	–	–	<0.001
<i>Patient variables</i>			
Age of patient, median (IQR), days	27 (11.3–42)	20 (5–42)	0.004
<i>Supervisor training in readiness assessment (2013–2014 only)</i>			
	Readiness assessed* (n=184)	Readiness not assessed* (n=101)	
Supervisor completed training	94 (51)	25 (25)	<0.001
Supervisor did not complete training	90 (49)	76 (75)	

\*Data are expressed as N (%).

†p Values calculated with  $\chi^2$  tests.

ED, emergency department; LP, infant lumbar puncture; NICU, neonatal intensive care unit; NS, not significant; PICU, paediatric intensive care unit.

**Table 2** Odds and potential confounders of successful LP

Factors	LP success rate	Unadjusted OR	Adjusted OR*
Completed assessment			
No	103 (35.0)	(Reference)	(Reference)
Yes	174 (40.2)	1.25 (0.92 to 1.70)	0.85 (0.60 to 1.20)
Specialty			
Paediatrics (n=600)	224 (37.3)	(Reference)	(Reference)
EM (n=74)	36 (48.6)	1.59 (0.98 to 2.58)	1.29 (0.74 to 2.24)
Med/Peds (n=44)	16 (36.4)	0.96 (0.51 to 1.81)	0.98 (0.48 to 1.99)
Location of LP			
Emergency department (n=334)	141 (42.2)	(Reference)	(Reference)
PICU (n=13)	7 (53.8)	1.60 (0.53 to 4.86)	2.23 (0.68 to 7.33)
NICU (n=164)	54 (32.9)	0.67 (0.45 to 0.99)†	0.97 (0.56 to 1.67)
Inpatient floor (n=189)	68 (36.0)	0.77 (0.53 to 1.11)	1.04 (0.64 to 1.69)
LP holder			
Technician (n=189)	88 (46.6)	(Reference)	(Reference)
Nurse (n=395)	143 (36.2)	0.65 (0.46 to 0.93)†	0.68 (0.43, 1.06)
Resident (n=64)	21 (32.8)	0.56 (0.31 to 1.02)	0.55 (0.27 to 1.15)
Fellow or attending (n=49)	16 (32.7)	0.56 (0.29 to 1.08)	0.50 (0.23 to 1.08)
Age of patient			
≤24 days (n=366)	132 (36.1)	(Reference)	(Reference)
>24 days (n=360)	145 (40.0)	1.20 (0.89 to 1.61)	0.94 (0.66 to 1.34)

Data are expressed as N (%).

\*Adjusted ORs calculated using mixed-effects, binary logistic regression with success as dependent variable and adjusting for assessment, specialty, location of LP, LP holder and age of patient (dichotomous) while accounting for intersite variances.

†p<0.05.

EM, emergency medicine; LP, infant lumbar puncture; Med/Peds, medicine and paediatrics; NICU, neonatal intensive care unit; PICU, paediatric intensive care unit.

**Table 3** LP outcomes and participation in readiness assessments (2012–2014)

Factors	Readiness assessed* (n=432)	Readiness not assessed* (n=294)	p Value†
LP outcomes			
Success on first attempt (primary outcome)	174 (40.0)	103 (35.0)	NS
Success on any attempt	203 (47.0)	127 (43.2)	NS
Number of attempts, median (IQR)	1 (1–2)	1 (1–2)	NS
Process outcomes			
Topical analgesia used	83 (19.2)	38 (12.9)	0.03
Early-styilet removal technique used	243 (56.3)	133 (45.2)	<0.001
Family member presence	112 (25.9)	61 (20.7)	NS
Attending supervision‡	195 (45.1)	86 (29.3)	<0.001

In all 726 patients who had an LP, we looked for additional factors associated with success including attending supervision, family member presence and analgesia use. No statistically significant associations were found.

\*Data are expressed as N (%).

†p Values calculated with  $\chi^2$  tests.

‡Intern LPs were otherwise supervised by senior residents or fellows.

LP, infant lumbar puncture.

A traumatic or unsuccessful LP in an infant quadruples the risk of hospitalisation and doubles the costs of the visit compared with those with a successful procedure.<sup>4</sup>

Clinical supervisors are obligated to determine whether a trainee is ready to perform a procedure. Workplace assessment instruments that focus on a trainee's readiness to perform concrete tasks can help supervisors make entrustment decisions.<sup>21 22</sup> For example, the Ottawa Surgical Competency Operating Room Evaluation (OSCORE) is one example of a behaviourally anchored, or 'independence'-rated ordinal scale that focuses on progression to competence.<sup>23</sup> Another benefit of assessing trainees in the workplace setting is that supervisors can then educationally intervene to help close knowledge gaps prior to patient encounters.<sup>21 22 24</sup> This emphasises the role of assessment as a learning experience in a well-integrated educational

experience.<sup>23</sup> The downside, however, is that supervisors tend to avoid rating trainees poorly, which threatens the reliability of a high-stakes assessment in the workplace.<sup>24</sup> In our study, this tendency to pass trainees may have impacted on supervisors' decisions to let trainees perform procedures.

Few studies have evaluated the implementation and impact of simulation-based readiness assessment in the clinical setting.<sup>2</sup> In our framework, applied across 35 institutions, we (1) trained learners using a common curriculum; (2) implemented a standard online faculty development programme; (3) increased workplace assessments for all first-time LP attempts and finally (4) implemented a centralised reporting mechanism allowing cross-institutional analysis of our impact. The timing of our study, concurrent with the Accreditation Council for Graduate Medical Education (ACGME) requirement to implement

milestones and competency assessments, helped us obtain support from programme directors to trial this novel assessment before entrusting residents to perform their first supervised LP.<sup>8 25–27</sup> Just-in-time readiness assessments have promised to provide information about a trainee's readiness at the point of care.<sup>14 28</sup> The real-world challenges we encountered of low uptake of faculty training and moderate overall participation have been confronted successfully in implementation studies of other skill-training paradigms.<sup>29</sup> We are hopeful that this LP training and assessment framework will lead to iterative refinement of our processes until LP success rates are uniformly high.

### Limitations

Owing to our large sample size, we were able to obtain a precise estimate of the success rate; however, we are limited in our ability to draw conclusions with regard to our selected comparisons. We thought it was unethical to randomise learners to preassessment versus none, in view of our prior validation study and the iterative small improvements we have seen in LP success rates. The historical comparison is prone to bias from both time and selection, while the concurrent comparison is also prone to selection bias. Without further demographic data, we cannot know whether these groups were fundamentally different in a way that might influence the main outcome measure. Other pragmatic limitations should be taken into account. The degree of realism for the simulator may have resulted in supervisor overestimates of resident readiness for procedures on actual patients. Reliance on self-report for procedural outcomes could miss procedures that were performed but not reported. Additionally, residents may have been inaccurate or dishonest in their reporting and we have no way to know if this tendency differed between our comparison groups. We did not collect data on how often a resident 'failed' the assessment or was denied the clinical procedure. This may have contributed to the low rate of participants who reported an LP; however, decreased exposure to procedures generally may also be a factor. Scant opportunities for clinical practice can also prompt supervisors to ignore unfavourable assessments (despite our best efforts to educate them on the risks of letting residents perform when not ready).<sup>2</sup> Heterogeneity of institutions and of individuals made it challenging to train all supervisors in the use of our readiness assessment and ensure consistency across programmes. However, by relying on sites to solve implementation problems locally, our study replicates general challenges seen with implementing any competency assessments in the workplace across multiple institutions.<sup>29</sup>

### CONCLUSIONS

Simulation-based readiness assessments were not associated with greater LP clinical success, but were related to several desirable behaviours including greater use of topical analgesia, use of the early stylet removal technique and more frequent attending supervision. We have shown that it is possible to put into place simulation-based readiness assessments across a large number of institutions, although further work is needed to improve implementation of the just-in-time training and assessment paradigm. Further patient outcome-oriented studies are needed to help determine the optimal method for assessing resident readiness for performing procedures on patients.

### Author affiliations

<sup>1</sup>Department of Pediatrics, Columbia University College of Physicians and Surgeons, New York, New York, USA

<sup>2</sup>Morgan Stanley Children's Hospital of New York Presbyterian, New York, New York, USA

<sup>3</sup>Department of Clinical Pediatrics, Keck School of Medicine of University of Southern California, Los Angeles, California, USA

<sup>4</sup>Children's Hospital of Los Angeles, Los Angeles, California, USA

<sup>5</sup>Department of Pediatrics and Emergency Medicine, Yale University School of Medicine, New Haven, Connecticut, USA

<sup>6</sup>Department of Pediatrics, Albert Einstein College of Medicine, New York, New York, USA

<sup>7</sup>Children's Hospital at Montefiore, Bronx, NY, New York, USA

<sup>8</sup>Department of Clinical Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

<sup>9</sup>The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

<sup>10</sup>Department of Pediatrics-Emergency Medicine and Medical Education, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

<sup>11</sup>Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA

<sup>12</sup>Department of Emergency Medicine, Stanford University School of Medicine, Stanford, California, USA

<sup>13</sup>Department of Pediatrics, Saint Louis University Health Sciences Center, Saint Louis, Missouri, USA

<sup>14</sup>SMM Cardinal Glennon Children's Medical Center, Saint Louis, Missouri, USA

<sup>15</sup>Department of Pediatrics, Emergency Medicine, Stony Brook Medicine, New York, USA

<sup>16</sup>Stony Brook Children's Hospital, New York, USA

<sup>17</sup>Yale School of Medicine, New Haven, Connecticut, USA

<sup>18</sup>Department of Emergency Medicine, New York University Langone Medical Center, New York, New York, USA

**Twitter** Follow David Kessler at @y2kessler

**Acknowledgements** The authors are grateful to the additional site directors who contributed to data collection and implementation at each site: Rebekah A Burns MD, Ann and Robert H Lurie Children's Hospital of Chicago; Daniel S Lemke MD, Baylor College of Medicine, Texas Children's Hospital; Jeffrey R Avner MD, Children's Hospital at Montefiore; Wendy Van Ittersum MD, FAAP, Cleveland Clinic Children's Hospital; Reena N Blanco, MD and Brian Costello MD, Emory University School of Medicine; Noel S Zuckerbraun MD, MPH and Brett L McAninch MD, University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh; Kevin J Ching MD and Nikhil B Shah MD, Weill Cornell Medical College. The authors would also like to acknowledge Keven Cabrera for his contributions to overall study coordination, data management and organisation; and to Vartan Pahalyants for his assistance with preparing this manuscript for submission.

**Contributors** DOK is primarily responsible for the conception and design of the study, drafting of the article and contributed to analysis and interpretation of data. TPC, MA, DMF, MEL, JT, MOL, JMG and DG were site directors primarily responsible for overseeing local data collection and study implementation. They all contributed meaningfully to study design, critically revising the manuscript and approved the final manuscript as submitted. MP and TW were primarily responsible for data analysis and interpretation. They were involved in aspects of study design, drafting and critical revision of the manuscript and approved the final version as submitted.

**Funding** All phases of this study were supported by a non-profit foundation grant (principal investigator: DOK) provided by the R Baby foundation.

**Competing interests** None declared.

**Ethics approval** Columbia University was primary, but all sites had their own local approval.

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

### REFERENCES

- 1 Nigrovic LE, Kuppermann N, Neuman MI. Risk factors for traumatic or unsuccessful lumbar punctures in children. *Ann Emerg Med* 2007;49:762–71.
- 2 Auerbach M, Chang TP, Reid J, et al. Are pediatric interns prepared to perform infant lumbar punctures? A multi-institutional descriptive study. *Pediatr Emerg Care* 2013;29:453–7.
- 3 Kessler D, Artega G, Ching K, et al. Interns' success with clinical procedures in infants after simulation training. *Pediatrics* 2013;131:e811–820.
- 4 Pingree EW, Kimia AA, Nigrovic LE. The effect of traumatic lumbar puncture on hospitalization rate for febrile infants 28 to 60 days of age. *Acad Emerg Med* 2015;22:240–3.
- 5 Shafer S, Rooney D, Schumacher R, et al. Lumbar punctures at an academic level 4 NICU: indications for a new curriculum. *Teach Learn Med* 2015;27:205–7.
- 6 Ten Cate OT. Nuts and bolts of entrustable professional activities. *J Grad Med Educ* 2013;5:157–8.
- 7 Tiyyagura G, Balmer D, Chaudoin L, et al. The greater good: how supervising physicians make entrustment decisions in the pediatric emergency department. *Acad Pediatr* 2014;14:597–602.
- 8 Schumacher DJ, Lewis KO, Burke AE, et al. The pediatrics milestones: initial evidence for their use as learning road maps for residents. *Acad Pediatr* 2013;13:40–7.

- 9 Kennedy TJ, Regehr G, Baker GR, *et al*. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med* 2008;83:S89–92.
- 10 Takayesu JK, Kulstad C, Wallenstein J, *et al*. Assessing patient care: summary of the breakout group on assessment of observable learner performance. *Acad Emerg Med* 2012;19:1379–89.
- 11 Scalse RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med* 2008;23(Suppl 1):46–9.
- 12 Braun C, Kessler D, Auerbach M, *et al*. Can residents assess other providers' infant lumbar puncture skills? Validity evidence for a global rating scale and subcomponent skills checklist. *Pediatr Emerg Care*. 2016 Oct 18. [Epub ahead of print].
- 13 Gerard JM, Kessler DO, Braun C, *et al*. Validation of Global Rating Scale and checklist instruments for the infant lumbar puncture procedure. *Simul Healthc* 2013;8:148–54.
- 14 Auerbach M, Fein DM, Chang TP, *et al*. The correlation of workplace simulation-based assessments with interns' infant lumbar puncture success: a prospective, multicenter, observational study. *Simul Healthc* 2016;11:126–33.
- 15 Pasternack JR, Dadiz R, McBeth R, *et al*. Qualitative study exploring implementation of a point-of-care competency-based lumbar puncture program across institutions. *Acad Pediatr* 2016;16:621–9.
- 16 Kessler D, Pusic M, Chang TP, *et al*. Impact of just-in-time and just-in-place simulation on intern success with infant lumbar puncture. *Pediatrics* 2015;135:e1237–1246.
- 17 Baxter AL, Fisher RG, Burke BL, *et al*. Local anesthetic and stylet styles: factors associated with resident lumbar puncture success. *Pediatrics* 2006;117:876–81.
- 18 Ellenby MS, Tegtmeyer K, Lai S, *et al*. Videos in clinical medicine. Lumbar puncture. *N Engl J Med* 2006;355:e12.
- 19 Auerbach M, Chang T, Krantz A, *et al*. Infant lumbar puncture: POISE pediatric procedure video. <https://www.mededportal.org/publication/8339> (accessed 14 Jan 2013).
- 20 Kessler DO, Auerbach M, Pusic M, *et al*. A randomized trial of simulation-based deliberate practice for infant lumbar puncture skills. *Simul Healthc* 2011;6:197–203.
- 21 Rekman J, Gofton W, Dudek N, *et al*. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med* 2016;91:186–90.
- 22 Embo M, Driessen E, Valcke M, *et al*. Integrating learning assessment and supervision in a competency framework for clinical workplace education. *Nurse Educ Today* 2015;35:341–6.
- 23 Gofton WT, Dudek NL, Wood TJ, *et al*. The Ottawa surgical competency operating room evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med* 2012;87:1401–7.
- 24 Schuwirth LW, Van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011;33:478–85.
- 25 Weller JM, Jones A, Merry AF, *et al*. Investigation of trainee and specialist reactions to the mini-clinical evaluation exercise in anaesthesia: implications for implementation. *Br J Anaesth* 2009;103:524–30.
- 26 Clinkard D, Moulton E, Holden M, *et al*. Assessment of lumbar puncture skill in experts and nonexperts using checklists and quantitative tracking of needle trajectories: implications for competency-based medical education. *Teach Learn Med* 2015;27:51–6.
- 27 Nasca TJ, Philibert I, Brigham T, *et al*. The next GME accreditation system—rationale and benefits. *N Engl J Med* 2012;366:1051–6.
- 28 Barsuk JH, McGaghie WC, Cohen ER, *et al*. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med* 2009;37:2697–701.
- 29 McGaghie WC, Barsuk JH, Cohen ER, *et al*. Dissemination of an innovative mastery learning curriculum grounded in implementation science principles: a case study. *Acad Med* 2015;90:1487–94.