BMJ Open

Proficiency-based progression training: an 'end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomised control study

Karthikeyan Kallidaikurichi Srinivasan,[1,2] Anthony Gallagher,[2] Niall O'Brien,[3,4] Vinod Sudir,[5] Nick Barrett,[6] Raymund O'Connor,[3,4] Francesca Holt,[3,4] Peter Lee,[2] Brian O'Donnell,[2] George Shorten[2]

To cite: Kallidaikurichi Srinivasan K, Gallagher A, O'Brien N, *et al*. Proficiency-based progression training: an 'end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomised control study. *BMJ Open* 2018;**8**:e020099. doi:10.1136/bmjopen-2017-020099

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-020099).

Received 13 October 2017
Revised 9 May 2018
Accepted 24 August 2018

For numbered affiliations see end of article.

Correspondence to
Dr Karthikeyan Kallidaikurichi Srinivasan; karthik@outlook.ie

ABSTRACT

Background Training procedural skills using proficiency-based progression (PBP) methodology has consistently resulted in error reduction. We hypothesised that implementation of metric-based PBP training and a valid assessment tool would decrease the failure rate of epidural analgesia during labour when compared to standard simulation-based training.

Methods Detailed, procedure-specific metrics for labour epidural catheter placement were developed based on carefully elicited expert input. Proficiency was defined using criteria derived from clinical performance of experienced practitioners. A PBP curriculum was developed to train medical personnel on these specific metrics and to eliminate errors in a simulation environment. Seventeen novice anaesthetic trainees were randomly allocated to undergo PBP training (Group P) or simulation only training (Group S). Following training, data from the first 10 labour epidurals performed by each participant were recorded. The primary outcome measure was epidural failure rate.

Results A total of 74 metrics were developed and validated. The inter-rater reliability (IRR) of the derived assessment tool was 0.88. Of 17 trainees recruited, eight were randomly allocated to group S and six to group P (three trainees did not complete the study). Data from 140 clinical procedures were collected. The incidence of epidural failure was reduced by 54% with PBP training (28.7% in Group S vs 13.3% in Group P, absolute risk reduction 15.4% with 95% CI 2% to 28.8%, p=0.04).

Conclusion Procedure-specific metrics developed for labour epidural catheter placement discriminated the performance of experts and novices with an IRR of 0.88. Proficiency-based progression training resulted in a lower incidence of epidural failure compared to simulation only training.

Trial registration number NCT02179879. NCT02185079; Post-results.

Strengths and limitations of this study

- First study of its kind comparing proficiency-based progression training versus simulation training looking at impact on patient outcome.
- The derivation of metrics, their validation and their application to training was carried out as part of one continuous 'end-to-end' process. This is the first report on the use of this methodology in its entirety, from procedure characterisation to meaningful clinical outcome.
- Single-centre study.
- Small sample size.

INTRODUCTION

Medical errors account for as many as 250 000 deaths in the USA every year.[1] A significant proportion of such errors (44% by one estimate) are related to procedural skills.[2] Expertise in certain procedural skills are associated with better patient outcomes.[3]

Simulation-based training[4] and assessment tools[5] have been developed to address the deficiencies in training and assessment of procedural skills. Although simulation training offers benefits in the training of procedural skills, evidence demonstrating transfer to the clinical setting or positive impact on patient outcomes is limited.[6–9] Assessment tools such as task-specific checklists (TSCL), global rating scales (GRS) and cumulative sum techniques[10 11] attempt to either (a) achieve better qualitative outcome (based on subjective assessment) or (b) rely on some form of self-reporting. The resulting limitation in objectivity undermines two critical characteristics of the assessment and training namely (i) inter-rater reliability (IRR) and (ii) facility to provide meaningful feedback to the learner.

This study aims to address these limitations by the use of a proficiency-based progression

(PBP) training methodology, based on unambiguously defined metrics. In PBP training, the learner is required to demonstrate a proficiency benchmark in procedure performance before progressing.[12 13]

We hypothesised that PBP training for epidural catheter placement for labour analgesia will result in better patient outcome (effective epidural analgesia) compared simulation training without PBP. This hypothesis was based on three assumptions. First, PBP is superior to conventional training for procedural skills.[14–19] Second, superior performance in a simulated setting will 'transfer' as superior performance in a clinical setting.[19] Third, superior procedural skills in the delivery suite will lead to improved patient outcomes (effective epidural analgesia).[3] Although these assumptions have been tested individually for various procedures, this is the first study in which the overall hypothesis, from metric definition to clinical outcome, has been tested.

METHODOLOGY

With Institutional ethical approval (September 2013) of the Cork Research Ethics Committee, and having obtained written informed consent from each participant, the study was conducted in three phases at Cork University Hospital and Cork University Maternity Hospital from September 2013 to September 2016.

Study phase 1: development of metrics

This phase was done between September 2013 and April 2014. A group of three experts (MW, BOD, PL) in lumbar epidural catheter placement were selected. An expert was defined as one who has performed more than 500 labour epidural catheter insertions in the preceding 5 year period. They attended five face-to-face meetings (each lasting for 120–180 min) facilitated by an investigator (KKS). The expert group identified and then defined procedure-specific metrics and errors. Metrics are units of observable behaviour which together constitute a stepwise description of a reference approach to a procedure, in this case, lumbar epidural catheter insertion for analgesia during labour. Errors are deviations from optimal procedure performance as described previously.[13]

Two video recordings of experts and two video recordings of novices performing epidural catheter insertion were acquired. Novices were defined as anaesthesia trainees with fewer than 2 years of experience and who had performed fewer than 50 epidural catheter insertions in total.[20] These videos were reviewed during metric development meetings.

Experts were requested to define each metric in the procedure objectively and explicitly. A metric could be either a step in the procedure or an error. Only metrics observable on reference videos were included. Assessment outcomes were defined dichotomously as 'yes or no' answers that is, that the metric as defined either had or had not occurred. For this particular procedure, all metrics were defined in terms of errors. Errors were categorised as critical (likely to cause actual patient harm) or noncritical (unlikely to cause actual harm but constituting a deviation from the defined or optimal approach). All discussions during the expert group meetings were audio recorded for review and future reference.

On completion of metric development, the experts independently scored two videos of labour epidural catheter placement (one by a novice and another by an independent expert) using the metrics-based assessment tool. Scores were compared and any reason for disagreement on rating between experts on specific metrics was discussed ('stress tested'). After refinement, a final list of metrics was approved by the expert group (online supplementary appendix 1).

Study phase 2: validation of metrics

This phase was done between April 2014 and December 2014. The metrics were then subjected to assessments for construct validity (a set of procedures for evaluating a testing instrument based on the degree to which the test terms identify the quality, ability or trait it was designed to measure) and concurrent validity (in which the relationship between the test scores and the scores on another instrument purporting to measure the same construct are related). We used GRS and TSCL[11] (online supplementary appendix 2 and online supplementary appendix 3) previously validated for epidural catheter placements to establish concurrent validity. Videos of eight experts and eight novices, each performing two lumbar epidural catheter placements for labour were video recorded following written informed consent both from the patient and the anaesthetist. A wearable camera-mounted glasses (1280*720 p, 30fps, Ottera technology Ltd, IE) was used to record the procedure from the first person perspective. Videos were entered into the study when they met the following criteria: (i) the entirety of the procedure was captured from the predefined start to end point (ii) the procedure was completed in full by the study participant (novice or expert) (iii) all defined metrics were observable on the video. For this procedure, since all the metrics were described as errors, only videos that showed all possible errors were included.

Eligible videos were then anonymised and submitted to two independent assessors (KH, OOS), blinded to the category of anaesthetist performing the procedure. The assessors had not participated in the development of the metrics. They were trained in evaluation of performance using the derived metrics:GRS and TSCL during a 3-hour training session. This training session involved a face-to-face meeting with the assessors in which detailed description of the metrics, TSCL and GRS were provided. The assessors then scored sample videos independently. Any discrepancies in their scores were discussed in detail. Training was provided until inter-rater reliability of 0.8 or above was achieved. On validation, the proficiency benchmark was based on (i) the absence of critical errors and (ii) the error count not exceeding the average expert-derived error count measured during this phase.

Study phase 3: impact of PBP training on patient outcome

This phase was done between January 2015 a September 2016. A prospective, randomised, single blind controlled trial was carried out at Cork University Maternity Hospital (CUMH). An investigator contacted eligible patients by telephone from a pool of registered pregnant patients of 32–38 weeks gestation, scheduled for delivery at CUMH. If the patients were agreeable to receive further information on the study, a detailed patient information sheet and consent forms were provided to them. The patients were also provided an online link to access the study information via- http://www.ucc.ie/en/assert/aboutthecentre/research/researchproject/. This was hosted on the University College Cork website. An investigator subsequently met with patients during an antenatal visit, addressed any questions and, if the patient was agreeable, written informed consent to participate was obtained.

Anaesthesia trainees of fewer than 2 years of experience and who had performed fewer than 50 epidural catheter placements were invited to participate. At the minimum, participating trainees had completed 10 labour epidural catheter placements and had been deemed capable of performing the procedure without a supervisor present. Participating trainees were randomly allocated to either group S (simulation training group) or group P (PBP group). Random allocation was done using computer-generated random numbers and allocations were enclosed in sealed envelopes by an investigator (GS) not involved directly in recruitment of the trainees. Participant enrolment and allocation was done by one of the study investigators (KKS, RO, FH). Patients and outcome assessors (midwives) were blinded to the study group. The following trainee information was collected using a questionnaire:

1. Experience in anaesthesia (total duration of experience in months).
2. Total number of epidurals performed to date (not limited to labour epidurals).
3. Total number of spinal anaesthetics performed to date.
4. Use of corrective eye glasses or contact lenses.
5. Dateof most recent epidural performed/attempted (whether labour or not).

All participating trainees were required to complete a set of psychometric and visuospatial tests to ensure homogeneity of the trainees, namely card-rotation test, cube comparison test, map planning test and Edinburgh handedness inventory.[20–22]

Labour epidural analgesia training was divided into two parts. During part one, trainees in both groups were given access to the same study material on labour epidural analgesia. An assessment test ('Select the best answer') based on the material provided was done within 2 weeks of provision of the material to trainees in group P only. Trainees in Group P were required to score a predefined pass percentage (80%) before they could proceed to the next phase of training. If the score was not met, additional time was given for the trainees to review the study material provided. No assessments were carried out at this stage of group S participants (consistent with standard training at our institution). Part two comprised a standardised workshop (didactic session and simulation training session) run for each participant within 4 weeks of receiving the study material. In group S, all participants received didactic teaching on the performance of labour epidural catheter placements (including all the metrics developed from the phase 1) followed by a simulation training session. Participants were instructed on the use of an epidural simulator (Manikin KKM43E, Cardiac services 2013, SISK Healthcare Group, UK) and allowed to practise (for up to 4 hours a day on two consecutive days) in the presence of, and with advice from, a clinical expert. The actual duration of simulator use was left to the discretion of the trainees. No assessment was done at the end of their simulation training session.

In group P, all participants received didactic teaching and the 74 metrics developed in phase 1 were described in detail using examples. A list of the metrics was provided to trainees in group P. Video recordings from phase 1 were used to illustrate how errors happen in 'real-life' clinical situations. Group P trainees received instruction on use of the same simulator. They were then required to practise, hands-on, each metric using the manikin. Focused feedback was given on how to avoid errors. Once the trainee had practised each metric, he or she demonstrated the procedure from the start to finish. Two assessors then used the study's validated assessment tool to independently score trainee performance on the simulator. Feedback on errors and critical errors identified during the procedure were provided. This process was repeated until the trainees attained a predetermined proficiency benchmark (as described earlier) on two consecutive procedures. Trainees in group P were not permitted to proceed to the next phase of the study until they had attained the proficiency benchmark.

Each participating trainee proceeded, within 2 weeks of completing the workshop, to perform clinical procedures in the labour ward at CUMH. In the event that 2 weeks elapsed before the opportunity to do so arose, the trainee underwent re-training. Data were obtained relating to the next 10 labour epidural catheter placements performed by each trainee.

An epidural was deemed to have failed if one or more of the following resulted: (i) accidental dural puncture; (ii) supervisor takeover; (iii) inadequate analgesia (presence of pain as perceived by the patient) during uterine contractions at 60 min from the time of epidural needle insertion (this was documented by the assigned midwife) or (iv) abandonment of the procedure. The midwife and the supervising consultant were unaware of the study group to which the participating anaesthetist belonged.

Secondary outcomes were the following: (1) Difference in learning curve between two groups. This was calculated by first looking the percentage of failures on the first epidural catheter placement in each group across the trainees. This was followed by second, third and so on till failure rates across trainees was calculated for all 10

epidural catheter placements across trainees. The groups were then compared to see if there was a difference between them. The other outcomes were (2) transfer of training from simulated learning environment to clinical environment and (3) patient satisfaction (within 1 week of delivery).

Transfer of training was assessed by evaluating first-person-perspective video recordings of epidural placements in a subset of study subjects. Patient permission was sought to video record epidural placement. Video recordings were acquired using wearable camera-mounted glasses. Videos were included when they met the criteria set out in the validation phase of the study. Eligible videos were then anonymised and assessed by two independent assessors (AR, PC), blinded to the identity and group assignment of the participant. The assessors who participated in this phase of the study were from different institutions, and were not involved in the development or validation of the metrics. The assessors were trained as described earlier until they were able to score performance with IRR > 0.8.

Patient satisfaction with the quality of their labour analgesia was assessed by telephone within 1 week of delivery. Patients were asked if they were satisfied with their analgesia during labour (answer – yes or no).

The following clinical data were also collected: accidental dural puncture, presence of supervisor, requirement to re-site the epidural catheter, type of delivery and analgesic efficacy of drugs administered via the epidural catheter if used for instrumental delivery or caesarean section.

Sample size calculation

Labour epidural failure rate (as defined above) for year 1 trainees is 25%, based on estimates from previous studies.[23] Based on the magnitude of effect of PBP training applied to other medical procedures[12 14] we anticipated a failure rate in interventional group to be 5%. Based on alpha=0.05 and beta=0.8, we estimated that a minimum sample size of 48 procedures per group was required. To allow for various contingencies, we recruited eight trainees per group, each of whom would perform 10 consecutive procedures, that is, a total of 80 procedures per group.

Statistics

Data were analysed for normality of distribution by visual inspection of Q-Q plot and by test of normality (Kolmogorov-Smirinov). Parametric data were summarised as mean and SD. Nonparametric data were summarised as median and inter-quaretile range.

Study phase 2: statistics

Each video was scored by two assessors independently. The average of the two scores was used as a final score (metrics, errors and critical errors) for the procedure. Analysis of variance was used to compare the error score between groups (experts and novices) and p value less than 0.05 was considered significant. For IRR, a proportion based on the number of agreements between assessors divided by total number of metrics (ie, proportionate agreement) was used. The merits of this approach have been extensively discussed elsewhere.[24] IRR>0.8 was considered acceptable.

Study phase 3: statistics

Student's t test was used to compare parametric continuous data. Non parametric data were compared using Mann-Whitney U test. χ^2 tests were used to compare categorical data. SPSS V.22 was used for statistical calculation (IBM, Armonk, New York, New York, USA).

Patient and public involvement

The primary research questions was to assess if the proficiency-based training programme led to reduction in failure rates of labour epidural analgesia. The study was conducted only in patients who requested epidural analgesia during labour. There was no obligation from the part of the patients to undergo any additional procedures. The timing of request for epidural analgesia was left to patients' preference. Patients were not directly involved in the design of the study. The results of the study will be submitted to the local ethical committee and updated on clinicaltrials.gov website.

RESULTS

Study phase 1: development of metrics

Seventy-four metrics were identified and defined, each of which represented either an error or a critical error. A total of 12 metrics (errors) were classified as critical (online supplementary appendix 1).

Study phase 2: validation of metrics

During the validation phase, 32 videos were acquired in total (16 expert, 16 novice) from which 13 expert videos and nine novice videos met the criteria for inclusion in the final analysis. Of the 10 videos that were excluded, one patient withdrew consent after the video recording had been obtained; during three procedures, the operator removed the recording device prior to completion of the procedure; in six videos, the camera did not capture all the procedural steps (figure 1).

The remaining 22 videos were anonymised and analysed. The construct validity data obtained using the different scales are summarised in table 1. The average number of errors, as measured by metrics, made by the expert group was less (16) than the equivalent in the trainee group (20) (p=0.02 based on ANOVA). The GRS scores (but not TSCL) demonstrated construct validity, that is, differentiate between expert and novice performance. It must be noted that not only was the IRR (calculated as detailed earlier) of metrics the highest (0.88) among the three, it also enabled differentiation of trainees and experts.

Study phase 3: impact of PBP training on patient outcome

A total of 17 trainees were recruited to participate in the study (figure 2, consort flow chart). Three participants

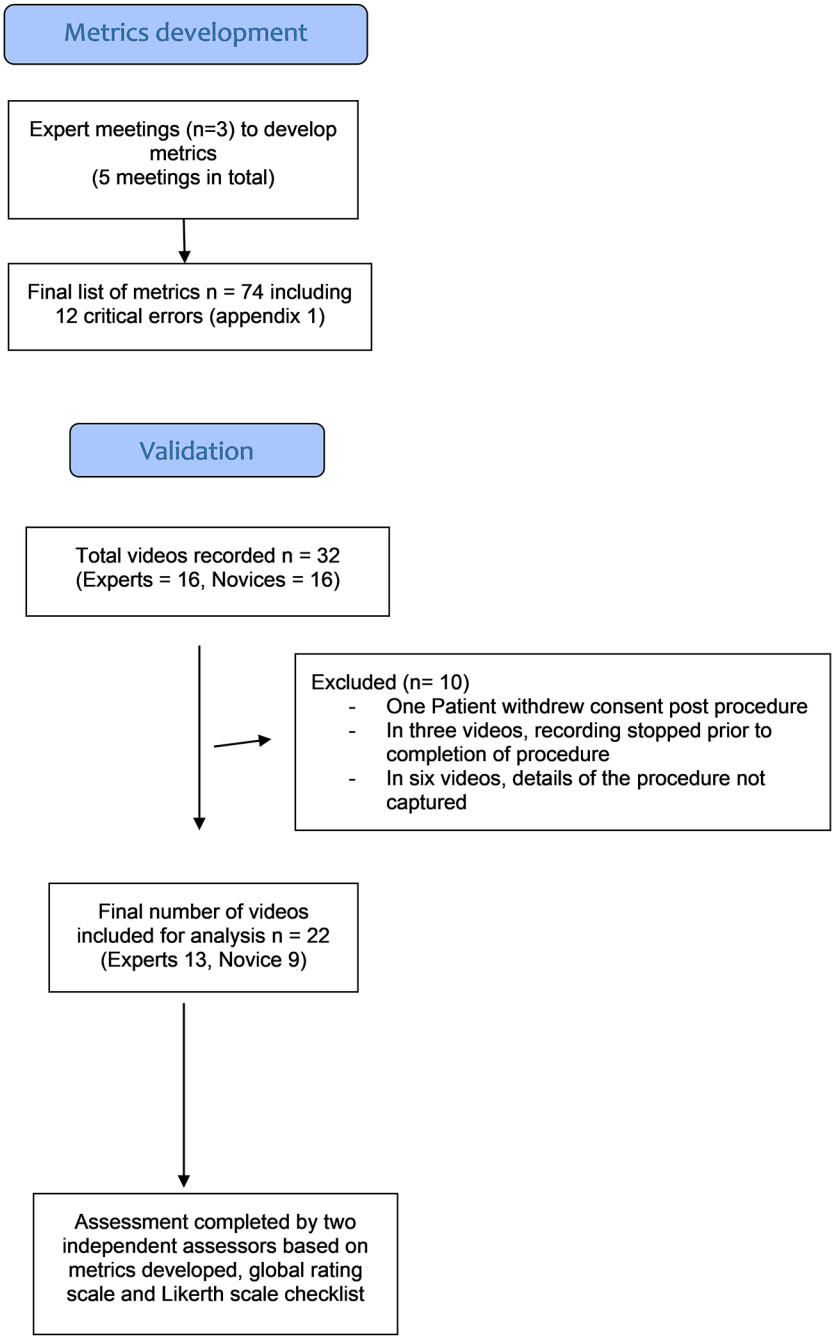

Figure 1 Outline of study phases 1 and 2.

were excluded from the study: one trainee from group S and two trainees from group P did not get an opportunity to perform labour epidural catheter placements within 2 weeks of completing training, and 14 participants completed the study, eight in group S and six in group P.

Trainee participant characteristics (table 2, online supplementary table 1) and participating parturient characteristic were similar in the two groups (table 2).

All trainees in group P achieved proficiency following three trials on a manikin. Trainees in group S and group P spent 97.8 min (SD 10.5) and 181.2 min (SD 12.5) respectively in completing the workshop (Student's t-test, p=0.0001).

Epidural catheter placement was performed on a total of 80 patients by participants in group S, and on 60 patients by participants in group P. The demographics, parity and type of delivery were similar between patient groups (online supplementary table 2).One procedure was abandoned and supervisor takeover occurred in eight procedures in group S. In group P, no procedures were abandoned and supervisor takeover occurred in two procedures (table 3).

The principal outcome— proportion of epidural failures— was greater in group S (23/80, 28.7%) than in group P (8/60, 13.3%) (p=0.04, Chi square test). The absolute risk reduction was 15.4% (CI 2% to 28.8%).

Table 1 Study phase 2 and 3: baseline parameters*

Baseline variables (study phase 2)	Novice		Expert		
Number of anaesthetists	5		6		
Age in years, median (minimum, maximum)	27 (24–32)		53 (44–57)		
Sex (M/F)	5/3		6/2		
Anaesthesia experience in years, median (minimum, maximum)	1		22.5 (12–25)		
Number of epidurals in past 5 years, median (minimum, maximum)	5 (2–12)		2000(1000–2500)		
	Group S		Group P		
Baseline variables (study phase 3)	**Median**	**IQR**	**Median**	**IQR**	**P values†***
Age	29	5.5	26	3	0.09
Experience in anaesthesia (in months)	17	7	18	6	0.92
Total number of epidurals performed prior to recruitment	16	17.5	10	13	0.29
Total number of spinal anaesthetics performed prior to recruitment	40	30	30	50	0.92
Most recent epidural performed prior to recruitment (days)	7	7	7	7	0.76

*All parametric data were summarised as mean and SD. All non-parametric data were summarised as median and IQR.
†Mann-Whitney U test.

The proportion of patients who experienced pain during uterine contraction at 60 min from the time of epidural needle insertion was also greater in group S (25%, 20/80) than in group P (10%, 6/60) (Chi square test P=0.03).

Twenty-one of the participating patients consented to undergo video recording (table 3). Of these, 17 were acquired in group P (trainee 1=10 videos, trainee 7=3 videos, trainee 12=2 videos, trainee 13=2 videos) and four in group S (trainee 6=3 videos, trainee 10=1 video). Eleven videos from group P and one video from group S met the criteria for inclusion. There were insufficient data from group S to perform an intergroup comparison, hence a proposed secondary outcome of the study (transfer of training) could not be assessed. Based on the limited information, group S (n=1 video) made a mean of 16.5 errors versus mean error of 4.3 (SD 1.8, CI 3.1 to 5.5) in group P (n=11 videos). The benchmark of proficiency was<16 errors (online supplementary table 3).

On comparison of the learning curves (online supplementary figure 1), the mean epidural failure rate in group S was 2.3 (SD 1.16) per 10 epidural catheter placements compared to a mean of 0.8 (SD 1.03) in group P (p<0.007, Mann-Whitney U test). Other epidural analgesia variables were similar in the two groups (online supplementary table 2).

DISCUSSION

Compared with simulation-only training (28.7%), PBP training was associated with a lower labour epidural failure rate (13.4%). This meant that we observed a 53% reduction in epidural failure rate in the PBP trained group.

Table 2 Study phase 2 : validity and inter-rater reliability of assessment scales*

	Trainees (n=16)		Experts (n=16)		
Assessment method	**Mean**	**SD**	**Mean**	**SD**	**P values†**
No of errors in metrics	20	1.59	16	4.6	0.02
Task-specific checklist score	46.9	2.3	48.8	2.7	0.23
Global rating scale score	21.7	2.7	31.6	1.4	<0.001

	IRR—trainees		IRR—experts		IRR—all procedures combined	
Assessment methods	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Metrics	0.86	0.02	0.88	0.06	0.88	0.05
Task-specific checklist	0.77	0.08	0.83	0.05	0.81	0.07
Global rating scale	0.15	0.12	0.46	0.14	0.33	0.2

*Student's t-test—two-tailed
†All parametric data were summarised as mean and SD. All nonparametric data were summarised as median and IQR.

Table 3 Study phase 3: labour analgesia variables

Variables	Group S (n=80)	Group P (n=60)	P values*
Accidental dural puncture, n (%)	0 (0)	0 (0)	–
Request for senior help, n (%)	10 (12.5)	6 (10)	0.79
Supervisor takeover, n (%)	8 (10)	2 (3.3)	0.19
Procedure abandoned, n (%)	1 (1.2)	0 (0)	0.57
Patient not comfortable at 60 min, n (%)	20 (25)	6 (10)	0.03
Reciting epidural at any stage, n (%)	6 (7.5)	5 (8.3)	0.55
Type of delivery, n (%)	Normal: 52 Instrumental: 15 Caesarean section: 13	Normal: 38 Instrumental: 12 Caesarean section: 10	0.98
Patient not satisfied with labour analgesia, n(%)	11 (13.7)	12 (20)	0.20

General anaesthesia for lower segment caesarean section: 0.
Spinal anaesthesia for lower segment caesarean section: two patients in group P.
*χ^2 test.

TSCL and GRS have been validated for assessment of epidural catheter insertion[11] and other procedural skills in anaesthesia.[25–27] The metrics-based assessment described in this study differs from these in two important ways. First, both TSCL and GRS use Likert scales for assessment. This necessarily introduces an element of subjectivity and limits their usefulness to providing detailed, specific feedback to the trainees.[28] Second, the use of Likert scales tends to decrease proportionate agreement, the form of IRR most relevant to high-stakes/risk procedural assessment. Correlation coefficients demonstrate association and not agreement.[24 29] For high-stakes assessment, a high level of inter-observer agreement is essential. Our results indicate that metrics-based assessment was satisfactory both in terms of discriminatory ability (establishing construct validity) and high IRR (score: 0.88).

The current study differs from others on PBP training[14 16] in two important ways: first the principal outcome was a meaningful clinical outcome (not simply performance quality) and second, the derivation of metrics, their validation and their application to training was carried out as part of one continuous 'end-to-end' process. This is the first report of use of this methodology in its entirety, from procedure characterisation to meaningful clinical outcome.

Epidural failure are reported as 8%–23%.[30–32] Thangamuthu *et al*[23] retrospectively reviewed 2169 epidurals performed in the UK over a 1-year period. Epidural failure was deemed to have occurred if one of the following was present: (i) inadequate analgesia reported at 45 min after epidural catheter placement; (ii) accidental dural puncture; (iii) abandonment of the procedure; (iv) the epidural catheter needed to be re-sited at any stage during labour; and/or (v) patient dissatisfaction with the analgesia provided at follow-up. Using the standard definition, the incidence of epidural failure rate was reported to be 26.8% in year 2 trainees and 17.4% in consultants. Patient satisfaction is subjective and can depend on factors other than adequate pain relief. Epidural catheter migration is known to occur either inwards (up to 13.7%) or outwards (up to 22.2%).[33] This might lead to deterioration in analgesia requiring re-siting of an appropriately sited epidural catheter and may not be a consequence of operator error. As our intention in this study was to objectively measure the initial failure rate associated with deficiencies in the procedure of catheter insertion, both (resiting of epidural catheters and patient satisfaction) were excluded from our definition of failure.

Limitations

The study has certain limitations. First, the study did not succeed in measuring one of its predefined secondary outcomes: procedure performance in the clinical setting. Videos obtained of trainees in group P demonstrated that the error rates were consistently and uniformly less than (ie, superior to) the predefined benchmark. Head-mounted cameras from which the captured video can be viewed live on a mobile phone are available and that may enable us to address this issue in the future.

Second, this was a single-centre study. The PBP training workshops were provided by authors who were involved with development of metrics from the development stage. Certain design elements of the study were intended to minimise the potential for institutional or investigator bias, namely: (i) the metric definition were required unambiguous descriptions of observable behaviours: in theory this should facilitate uniformity of feedback for a given performance and thus external validity; (ii) proficiency benchmark criteria were unambiguous; (iii) none of the assessors participated in metric development; although they were from different institutions good IRR was demonstrated for assessments; (iv) no attempt was made to measure 'skill of optimising epidural analgesia' for example, timing, dose and selection of agents for topups. This ultimately will influence overall quality of analgesia during labour; our focus was on initial achievement of satisfactory analgesia.

Third, the small sample size of our study necessitates care in interpreting the results. The overall difference

in failure rates between groups is 15.4%. Although this appears substantial, the 95% CI was 2% to 28.8% for difference in proportion of failures between groups.

Finally, the PBP group spent on average 83 min more than the simulation training group during the workshop. While one might argue that this might explain the difference in outcomes, it must be noted that both groups had access to similar study material preworkshop and had access to the same manikin under the guidance of the same instructors. Systematic reviews on simulation training have failed to identify any difference between simulation training and nonsimulation training in terms of patient outcomes in spite of additional time spent on teaching.[4 9] Hence we believe that the increase in training time in itself may not sufficiently explain the difference between the two groups.

The aim of the study was to investigate if the methodology will lead to better patient outcomes. We acknowledge the resource intensiveness of this intervention but this is a proof -of-concept study. Further research could be done on mitigating/optimisation of the resource intensiveness of this methodology.

Summary

Procedure-specific metrics developed for labour epidural catheter placement discriminated the performance of experts and novices with IRR of 0.88. PBP training with simulation based on these metrics the decreased epidural failure rates by 53% when compared with that of trainees who underwent 'simulation only' training. We have described an 'end-to-end' methodology, which may enable improvement in patient outcome for specific medical procedures.

Author affiliations
[1]Department of Anaesthesia, Adelaide and Meath Hospital, Dublin, Ireland
[2]The ASSERT Centre, University College Cork, Cork, Ireland
[3]The Department of Anaesthesia, Cork University Hospital, Cork, Ireland
[4]The Department of Anaesthesia, University College Cork, Galway, Ireland
[5]Department of Anaesthesia, SAOLTA University Healthcare Group, Galway, Ireland
[6]Department of Anaesthesia, Waterford University Hospital, Waterford, Ireland

Acknowledgements (1) Dr Mary Walsh, FFARCSI, Consultant Anaesthetist, Cork University Hospital, Cork for the valuable time and support for this project. (2) Professor Brendan Bunting, Professor of Psychology, Ulster University, Northern Ireland for his valuable inputs on the sample size calculation. (3) Olivia mason (Biostatistician, Centre for Support and Training in Analysis and Research, UCD, Ireland) for her inputs in to the overall approach to statistics.

Contributors The following were the criteria for author contributions: 1. Contribution to conception and design- KKS, AG, PL, BO'D, GS;. 2. Acquisition of data, or analysis and interpretation of data: KKS, AG, NO'B, VS, NB, RO'C, FH, PL, BO'D, GS; 3. Drafting the article or revising it critically for important intellectual content: KKS, AG, NO'B, VS, NB, RO'C, FH, PL, BO'D, GS; 4. Final approval of the version to be published: KKS, AG, NO'B, VS, NB, RO'C, FH, PL, BO'D, GS; and 5. Agreement to be accountable for all aspects of the work thereby ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: KKS, AG, NO'B, VS, NB, RO'C, FH, PL, BO'D, GS.

Funding The ASSERT (Application of Science to Simulation based Education and Research on Training) Centre (University College Cork, Ireland) funded the study by providing resources namely manikin for simulation training and research assistant to help with the consent process and data collection. All the work by the authors, inputs from the experts and participation of trainees was voluntary. No aid (monetary or otherwise) was provided in any way. Cork University Hospital provided the infrastructure for the study to take place. 6. Contribution to conception and design, acquisition of data or analysis and interpretation of data; 7. Drafting the article or revising it critically for important intellectual content; 8. Final approval of the version to be published; and 9. Agreement to be accountable for all aspects of the work thereby ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests None declared.

Patient consent Not required.

Ethics approval Clinical Research ethics committee, University College Cork.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement 1. Deidentified participant data (including data dictionaries) will be shared. 2. Data will be available for a period of up to 3 years following publication of the trial. 3. Data request from institutions or authorised personnel from institutions for research purposes will be facilitated.

REFERENCES

1. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ* 2016;353:i2139.
2. Institute of Medicine, Committee on Quality of Health Care in America. *To err is human: building a safer health system*. Washington, DC: The National Academies Press, 2000.
3. Birkmeyer JD, Finks JF, O'Reilly A, *et al*. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434–42.
4. Lorello GR, Cook DA, Johnson RL, *et al*. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. *Br J Anaesth* 2014;112:231–45.
5. Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. *Br J Anaesth* 2009;103:472–83.
6. Ross AJ, Kodate N, Anderson JE, *et al*. Review of simulation studies in anaesthesia journals, 2001-2010: mapping and content analysis. *Br J Anaesth* 2012;109:99–109.
7. Stefanidis D, Sevdalis N, Paige J, *et al*. Simulation in surgery: what's needed next? *Ann Surg* 2015;261:846–53.
8. Okuda Y, Bryson EO, DeMaria S, *et al*. The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med* 2009;76:330–43.
9. Zendejas B, Brydges R, Wang AT, *et al*. Patient outcomes in simulation-based medical education: a systematic review. *J Gen Intern Med* 2013;28:1078–89.
10. de Oliveira Filho GR. The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method. *Anesth Analg* 2002;95:411–6.
11. Friedman Z, Katznelson R, Devito I, *et al*. Objective assessment of manual skills and proficiency in performing epidural anesthesia–video-assisted validation. *Reg Anesth Pain Med* 2006;31:304–10.
12. Gallagher AG, Seymour NE, Jordan-Black JA, *et al*. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. *Ann Surg* 2013;257:1025–31.
13. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. *Ulster Med J* 2012;81:107–13.
14. Ahlberg G, Enochsson L, Gallagher AG, *et al*. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg* 2007;193:797–804.
15. Cates CU, Lönn L, Gallagher AG. Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators. *BMJ Simulation and Technology Enhanced Learning* 2016;2:1–5.
16. Angelo RL, Ryu RK, Pedowitz RA, *et al*. A proficiency-based progression training curriculum coupled with a model simulator

results in the acquisition of a superior arthroscopic bankart skill set. *Arthroscopy* 2015;31:1854–71.
17. Pedowitz RA, Nicandri GT, Angelo RL, *et al*. Objective assessment of knot-tying proficiency with the fundamentals of arthroscopic surgery training program workstation and knot tester. *Arthroscopy* 2015;31:1872–9.
18. Van Sickle KR, Ritter EM, Baghai M, *et al*. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg* 2008;207:560–8.
19. Seymour NE, Gallagher AG, Roman SA, *et al*. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236:458–63.
20. Kopacz DJ, Neal JM, Pollock JE. The regional anesthesia "learning curve". What is the minimum number of epidural and spinal blocks to reach consistency? *Reg Anesth* 1996;21:182–90.
21. Gallagher AG, Smith CD, Bowers SP, *et al*. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. *J Am Coll Surg* 2003;197:479–88.
22. Tansley P, Kakar S, Withey S, *et al*. Visuospatial and technical ability in the selection and assessment of higher surgical trainees in the London deanery. *Ann R Coll Surg Engl* 2007;89:591–5.
23. Thangamuthu A, Russell IF, Purva M. Epidural failure rate using a standardised definition. *Int J Obstet Anesth* 2013;22:310–5.
24. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc* 2003;17:1525–9.
25. Naik VN, Perlas A, Chandra DB, *et al*. An assessment tool for brachial plexus regional anesthesia performance: establishing construct validity and reliability. *Reg Anesth Pain Med* 2007;32:41–5.
26. Chin KJ, Tse C, Chan V, *et al*. Hand motion analysis using the imperial college surgical assessment device: validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Reg Anesth Pain Med* 2011;36:213–9.
27. Burckett-St Laurent DA, Niazi AU, Cunningham MS, *et al*. A valid and reliable assessment tool for remote simulation-based ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2014;39:496–501.
28. Gallagher AG, O'Sullivan GC, Leonard G, *et al*. Objective structured assessment of technical skills and checklist scales reliability compared for high stakes assessments. *ANZ J Surg* 2014;84.
29. Chuan A, Graham PL, Wong DM, *et al*. Design and validation of the regional anaesthesia procedural skills assessment tool. *Anaesthesia* 2015;70:1401–11.
30. Roberts RB. The occurrence of unblocked segments during continuous lumbar epidural analgesia for pain relief in labour. *Br J Anaesth* 1972;44:628.
31. Michael S, Richmond MN, Birks RJ. A comparison between open-end (single hole) and closed-end (three lateral holes) epidural catheters. Complications and quality of sensory blockade. *Anaesthesia* 1989;44:578–80.
32. Pan PH, Bogard TD, Owen MD. Incidence and characteristics of failures in obstetric neuraxial analgesia and anesthesia: a retrospective analysis of 19,259 deliveries. *Int J Obstet Anesth* 2004;13:227–33.
33. Bishton IM, Martin PH, Vernon JM, *et al*. Factors influencing epidural catheter migration. *Anaesthesia* 1992;47:610–2.