Development of performance and error metrics for ultrasound-guided axillary brachial plexus block

Osman M Ahmed 1
Brian D O’Donnell 1,2
Anthony G Gallagher 2
George D Shorten 1,2
1Department of Anaesthesia and Intensive Care, University College Cork and Cork University Hospital, 2The ASSERT Center, University College Cork, Cork, Ireland

Purpose: Change in the landscape of medical education coupled with a paradigm shift toward outcome-based training mandates the trainee to demonstrate specific predefined performance benchmarks in order to progress through training. A valid and reliable assessment tool is a prerequisite for this process. The objective of this study was to characterize ultrasound-guided axillary brachial plexus block to develop performance and error metrics and to verify face and content validity using a modified Delphi method.

Methods: A metric group (MG) was established, which comprised three expert regional anesthesiologists, an experimental psychologist and a trained facilitator. The MG deconstructed ultrasound-guided axillary brachial plexus block to identify and define performance and error metrics. Experts reviewed five video recordings of the procedure performed by anesthesiologists with different levels of expertise to aid task deconstruction. Subsequently, the MG subjected the metrics to “stress testing”, a process to ascertain the extent to which the performance and error metrics could be scored objectively, either occurring or not occurring with a high degree of reliability. Ten experienced regional anesthesiologists used a modified Delphi method to reach consensus on the metrics.

Results: Fifty-four performance metrics, organized in six procedural phases and characterizing ultrasound-guided axillary brachial plexus block and 32 error metrics (nine categorized as critical) were identified and defined. Based on the Delphi panel consensus, one performance metric was modified, six deleted and three added.

Conclusion: In this study, we characterized ultrasound-guided axillary brachial plexus block to develop performance and error metrics as a prerequisite for outcome-based training and assessment. Delphi consensus verified face and content validity.

Keywords: task analysis, metrics, validation, training, assessment

Introduction
Ultrasound-guided regional anesthesia (UGRA) has evolved rapidly to become a fundamental skill for anesthesiologists. UGRA comprises a group of procedures that facilitate provision of anesthesia and analgesia for different surgical procedures (e.g., upper and lower limbs surgery) either alone or in combination with general anesthesia. It is usually performed by anesthesiologists with advanced level of expertise and requires mastery of a wide range of technical and nontechnical skills. Widespread adoption of UGRA by anesthesiologists has created a requirement to teach and learn UGRA-related skills effectively. Training bodies and groups have endorsed a number of structured training programs to aid skills acquisition during performance of UGRA.
the American and European Societies of Regional Anesthesia. The current paradigm shift toward competence-based training has led to a demand for objective, valid and reliable assessment tools for procedural skills in general. However, characterization and development of assessment tools for anesthesia-related procedural skills are less advanced compared to their surgical counterparts. It has been argued that assessment motivates learning and facilitates provision of structured feedback leading to improved procedural skills training. Assessment of competence could be greatly improved when direct observation with checklists is used. Disciplines such as surgery have successfully applied a metric-driven proficiency-based progression approach to surgical technical skills training. This approach (in contrast to generic or Objective Structured Assessment of Technical Skills approach) requires detailed characterization of each specific procedure by unambiguous definition of specific events or “metrics”.

Procedures such as UGRA may be more effectively taught if mastered initially as component parts, subsequently assimilated into an integrated performance of the complete procedure. Collaboration with experienced operators who are proficient in performance of the procedure permits identification of 1) what should be done (the essential steps), 2) how it should be done (effective techniques) and 3) what should not be done (errors). This expert-driven procedure characterization must then be subjected to a validation process, for example, face and content validity.

The Delphi method is a process that utilizes communication between experts in a field in order to provide a feedback on a given topic. The Delphi method uses an iterative process for progressing toward a desired result by means of repeated cycles of deliberation.

The objectives of this study were
1. to characterize procedure performance and develop metrics for ultrasound-guided axillary brachial plexus block using qualitative techniques;
2. to verify face and content validity of the metrics using a modified Delphi method.

### Methods

**Procedure characterization and metrics generation by metric group**

With approval of the Clinical Research Ethics Committee of the Cork Teaching Hospitals and the expressed consent of all participants, a metric group (MG) composed of three experts (BOD, FL, PM) was established. Experts were defined as anesthesiologists who 1) had undergone formal higher subspecialty training in regional anesthesia, 2) perform ultrasound-guided axillary brachial plexus blocks regularly and 3) currently teach UGRA to trainee anesthesiologists and regional anesthesia fellows. All the three members of the MG are current practicing consultant anesthesiologists and satisfied the aforementioned criteria of an expert. The MG also included an experimental psychologist (AGG) and a trained facilitator (OMA) who moderated the interactions of the MG to achieve the study objectives.

The investigators defined start and end points for the procedure prior to commencement of the MG meetings. The MG undertook a “task analysis” and deconstruction process, detailed previously by Gallagher et al., to identify units of performance that are integral to skilled performance of ultrasound-guided axillary brachial plexus block. During six face-to-face meetings (each of 120–150 minutes duration) over a 3-month period, performance metrics were identified and then unambiguously defined. The meeting moderator focused on the group interaction to ensure that the definitions derived would enable an independent assessor to subsequently use them with a high degree of reliability. The aims were to 1) make sure that these performance metrics should capture the core elements of procedure performance and 2) create unambiguous operational definitions for each performance metric. Experts included the specific steps and general order (when relevant) in which the procedure should be performed. Procedural phases were specified for groups of related performance metrics. The MG also identified and unambiguously defined potential “error metrics” or actions that deviate from optimal performance. Two types of error metrics are identified within each of the procedure phases. Noncritical errors are behaviors that deviate from optimal performance and might compromise procedure success but do not compromise patient safety. Critical errors are behaviors that deviate from optimal performance and compromise patient safety and might cause actual patient harm. Experts were specifically instructed to characterize a “reference” procedure (i.e., a straightforward procedure) and not an unusual or complex procedure. It was emphasized throughout that the MG should seek to reach consensus, not necessarily agreement between the three experts on the performance and error metrics. Five video recordings of complete live procedures performed by anesthesiologists with different levels of expertise (consultant and trainee anesthesiologists) were reviewed in detail during MG meetings to aid task deconstruction.
Videography of ultrasound-guided axillary brachial plexus block

An investigator acquired a video recording of the entire procedure according to specifically predefined start and end points. Written informed consent was obtained from each patient and participating anesthesiologist. Patients were assessed to be ASA I–III scheduled to undergo upper limb orthopedic or plastic surgery procedure, and for whom an axillary brachial plexus block was planned, by the responsible consultant anesthesiologist as part of their perioperative management.

Two cameras were used to record the procedure: first-person perspective imaging was obtained using a head-mounted camera (GoPro™ Hero GP 1049 [Patented USA] GoPro Inc., San Mateo, CA, USA) placed on operator’s head; third-person perspective imaging was obtained by the investigator using a handheld camera (Sony HX5V HD; Sony Corporation of America, New York, NY, USA). Concurrent ultrasound video images were recorded using the DVR feature of the Ultrasonix tablet (Sonix Design Centre BK Ultrasound, Richmond, BC, Canada). All video recordings were edited using iMovie (Apple iMovie, version 10.0.5; Apple Computer Inc., Cupertino, CA, USA) to present the procedure, from first- or third-person perspective as appropriate, together with the ultrasound output on one screen. This was achieved using “picture-in-picture” mode (Figure 1). The ultrasound feed was displayed on the right side of the screen, taking care not to obscure the view of the operator’s hands. Both the procedure and ultrasound video recordings were time synchronized. The identities of patient and participant were masked. These video recordings were reviewed by the MG to facilitate their effort to ensure that the defined metrics were both observable and valid.

Performance and error metrics reliability assessment “stress testing” by MG

Following the “task analysis” phase, the MG examined the performance and error metrics for reliability testing “Stress Testing”. Each of the three members, independently, reviewed and scored two video recordings of the procedure that were acquired during the metrics development stage, in a blinded fashion. After viewing and scoring the recordings, scoring patterns from all the three experts were compared and discussed in a joint and final (seventh) MG meeting.

Examination of face and content validity using a modified Delphi method

Once the MG members were satisfied that all procedure phases were appropriately characterized, the performance and error metrics were subjected to face and content validity testing using a modified Delphi panel method.

Ten consultant anesthesiologists each of whom was a member of Irish Society of Regional Anaesthesia and who are also experts in regional anesthesia (as per the aforementioned criteria) participated in this process. Members of the MG, who had originally developed the metrics, did not participate in the Delphi panel.

A brief overview of the study and meeting objectives were presented by one of the investigators. Background information regarding task analysis, procedure characterization and the specific objectives of the current Delphi panel were presented. Each phase of the procedure and the associated performance and error metrics were presented. As part of this briefing, it was emphasized that the designated “reference procedure” was not intended to reflect the exact technique employed by an individual expert practitioner, but to show that the performance metrics presented were intended to accurately capture the essential and key components of the procedure. The specific questions asked were:

1) Do these metrics represent a reasonable characterization of the procedure? 2) Do any of the metrics represent an incorrect approach? Members of the Delphi panel were also asked if a performance or error metric definition should be modified or if a metric should be added or deleted. After each of the procedure phases was presented, attendees were given time to review individual metrics in detail. After each metric definition was reviewed, panel members voted on whether or not the metric was acceptable as written. An affirmative vote by a panel member indicated that the metric definition presented was accurate and acceptable as written. If consensus could not be achieved, then modifications as suggested
by the panel were employed and members re-voted until consensus was achieved.

Results

The performance metrics developed (categorized in six procedural phases) are summarized in Table 1. Error metrics (noncritical and critical) are summarized in Table 2. During the final (seventh) MG meeting, the three members subjected the performance and error metrics to “stress testing”. The outcome was a final modification of two metrics relating to maintaining asepsis, which were condensed to appear after section E (metrics 42, 43; Table 1) rather than after each of the preceding sections.

The Delphi panel meeting was held on May 21, 2015, during the Irish Congress of Anesthesia (ICA) in Dublin. The meeting lasted for 180 minutes. Three phases (B, F and E) attained consensus with a single round of voting and no modifications were made. Three phases (A, C and D) required modifications and subsequently achieved consensus after a second round of voting. The amendments that were derived from the Delphi panel meeting are as follows:

Phase A: Four members voted against performance metric 6 (Table 1) so that it was modified to “sedation to be administered as clinically indicated” (not on patient request) and subsequently attained consensus in the second round of voting.

Phase C: Six members voted to delete two performance metrics; one about wearing a facemask and the other about using povidone–iodine (Betadine) as an antiseptic; the consensus was to delete these. Members also voted to delete metric 21 (Table 1) as not needed for single shot block, but it was subsequently retained following the second round of deliberation; members discussed and subsequently voted on adding performance metric 22 (Table 1).

Phase D: All members agreed that performance metrics related to operator verbalizing their interpretation of ultrasonography scan (four performance metrics) would introduce a distraction. Following further discussion, members reached a consensus to evaluate that phase of the procedure based on two specific “outcome” performance metrics (metrics 28, 29; Table 1).

Discussion

In this study, we characterized skilled performance of ultrasound-guided axillary brachial plexus block. Face and content validity of the outcome of the task characterization (metrics) had been examined and confirmed using a modified

Table 1 Six phases of the procedure characterization and abbreviated summary of the definitions of the 54 performance metrics developed by the MG and validated by the Delphi panel

<table>
<thead>
<tr>
<th>Phase (Abbreviation)</th>
<th>Description</th>
</tr>
</thead>
</table>
| A. Initial patient interaction | 1. Place patient on a suitable trolley or bed  
2. Establish conversation with the patient  
3. Prepare safe surgical checklist  
4. Establish an intravenous cannula  
5. Establish monitoring  
6. Give sedative medication  
7. Administer supplemental oxygen |
| B. Procedure ergonomics | 8. Adjust equipment within workplace  
9. Remove clothing from arm  
10. Expose patients arm  
11. Place arm in block position |
| C. Aseptic preparation (patient, operator, equipment) | 12. Ensure operator arms are bare  
13. Decontaminate hands  
14. Wear sterile gloves  
15. Prepare block tray  
16. Prepare a high-frequency ultrasound probe  
17. Prepare local anesthetics solution  
18. Confirm name and expiry date of local anesthetics  
19. Flush block needle  
20. Decontaminate skin of the axilla  
21. Establish sterile field by using sterile drapes  
22. Fix sterile drapes |
| D. Ultrasonography | 23. Check sidedness of the ultrasound probe  
24. Place probe on the axilla  
25. Hold probe in a tripod grip  
26. Confirm identity of axillary artery  
27. Immobilize the probe  
28. Acquire an optimum ultrasound image  
29. Verbalize the main components of the optimum ultrasound image |
| E. Needle guidance and injectate placement | 30. Inform the patient of needle insertion  
31. Advance block needle adjacent to ultrasound probe  
32. Keep needle visible when advancing  
33. Reestablish needle visibility during forward motion  
34. Withdraw needle toward skin to reestablish visibility  
35. Maintain communication with patient  
36. Aspirate before injection of local anesthetics  
37. Maintain patient veins during injection  
38. Encourage patient to report pain on injection  
39. Use test dose of local anesthetics  
40. Place full block dose  
41. Place sufficient amount of local anesthetics  
42. Maintain sterility of procedure field  
43. Correct breaches of sterility if required |
| F. Block clinical assessment | 44. Remove cast or splint from the blocked arm  
45. Establish presence or absence of pain  
46. Assess motor power along distribution of median nerve  
47. Assess motor power along distribution of ulnar nerve  
48. Assess motor power along distribution of radial nerve proximal fibers  
49. Assess motor power along distribution of radial nerve distal fibers  
50. Assess motor power along distribution of ulnar nerve  
51. Assess sensory function along distribution of median nerve  
52. Assess sensory function along distribution of ulnar nerve  
53. Assess sensory function along distribution of radial nerve  
54. Assess sensory function along distribution of musculocutaneous nerve |

Abbreviation: MG, metric group.
Delphi method. In addition to the specific new performance and error metrics developed, the study describes in detail the application of this methodology to an image-guided procedure with patient needle interaction for the first time.

Procedural skills play an important role in anesthetic practice. Evidence-based training and assessment of procedural skills are necessary as training moves from an apprenticeship-based model to one of competency-based training. Our results are consistent with previous studies. Angelo et al adopted a similar methodology to characterize task and develop performance and error metrics for arthroscopic shoulder surgery. Subsequently, they successfully incorporated the metrics into a training curriculum in simulated learning environments. For ultrasound-guided nerve blocks, checklists and global rating scales (GRS) have been developed and examined for validity. Both Cheung et al and Chuan et al developed a “one-size-fits-all” assessment tool for UGRA to include checklist and GRS and validated the tool by using a modified Delphi method. GRS as used in these two studies permits a degree of subjectivity, which may influence the assessment outcome. This subjectivity in turn hampers interrater reliability, an essential component of a valid assessment tool. The methodology used in this study requires detailed characterization of the procedure by generating unambiguously defined performance and error metrics and their subsequent examination for face and content validity. This approach has been quantitatively shown to have greater assessment reliability when compared to Likert scale assessments used with GRS. One of the strengths of this study is that each performance and error metric represents an observable behavior that is precisely defined within the context of ultrasound-guided axillary brachial plexus block.

This study provided one prerequisite element of an evidence-based approach to training ultrasound-guided axillary brachial plexus block, namely, the development of an assessment tool. More work will be necessary to integrate this into an overall proficiency-based training program and to prospectively examine the resulting performance effect versus the best alternative currently available. The proficiency-based progression approach mandates that the trainee be able to demonstrate the ability to meet specific predefined benchmarks to progress in training. These benchmarks must have specific, clear, objective and fair standards of performance. The metrics that we described in this study will be essential in defining these standards. The importance of this study is that it provides data and a framework for the development of such a standard (although this was beyond the scope of this study).

It is also important for the task analysis to identify behaviors that deviate from optimal performance, i.e., “error metrics”. It has been proven that error-focused checklists are superior to conventional checklists in terms of evaluating procedural incompetence. Any training program based on the performance and error metrics generated in this study will need to be evaluated (versus the best available alternative) in terms of any influence in incidence of defined error metrics in both simulated and clinical settings.

Various forms of checklists have been applied to assessment of procedural and nontechnical skills. What differentiates the approach described in this study is the rigor and methodological detail applied to generating, defining and
stress testing of each performance and error metric. We believe that the resulting unambiguous definitions effectively “capture” the critical elements of performance and underpin the efficacy of subsequently developed training programs. In the past (when applied to surgical procedures), the training efficacy of such programs has been invariable and substantial.

The study has potential weaknesses. The three experts in the MG are very experienced in the performance of ultrasound-guided axillary brachial plexus block. We acknowledge that three (experts) is a small number of sources from which to elicit details of a reference procedure, which is performed in many different settings around the world. The fact that all the three experts work in a single center raises the possibility that their views represent or contain some institutional bias. These limitations are mitigated by the use of the Delphi process, proven to be an excellent and reliable method to test expert view for consistency with that of a wider practicing community. Another important limitation of this study relates to the definition of metrics as observable behaviors when a substantial part of the procedure entails cognitive elements that are not “observable”. Much interpretation of the initial ultrasound scan is based on cognitive function, which is not directly captured as observable behavior. We judged that prompting the operator to verbalize their interpretation or reasoning during the ultrasound image acquisition would introduce a distraction and an unwarranted demand on cognitive resources and may cause an operator to deviate from his/her intended or normal practice. Based on the Delphi panel discussion, it was decided to evaluate that phase of the procedure based on two specific “outcome” metrics (28, 29; Table 1) to minimize distraction caused by prompting.

We also acknowledge the need to examine the set of performance and error metrics described here for construct validity and reliability. This has been investigated recently using a prospective observational design. Future studies are required to evaluate the effectiveness of new training programs based on these metrics.

Conclusion
In summary, we applied a methodology (successful in surgical technical skills training) to characterize performance of ultrasound-guided axillary brachial plexus block. The performance and error metrics presented here have face and content validity and may be used to objectively assess trainee skills acquisition and facilitate feedback during training. Integration of these into proficiency-based progression training program will require definition of performance benchmarks and prospective examination of the novel training program versus the best alternative.

Acknowledgment
The authors thank Dr Frank Loughnane and Dr Padraig Mahon for their contributions during the metrics development. They were members of the MG and participated in the identification and definition of metrics and errors.

Author contributions
Conception and design of study: OMA, BDO, AGG, and GDS; data acquisition, analysis and interpretation: OMA, BDO, AGG, and GDS; and drafting, revising and approval of the manuscript: OMA, BDO, AGG, and GDS. All the authors agreed to be accountable to all aspects of publication process.

Disclosure
The authors report no conflicts of interest in this work.

References
Development of performance metrics


