COMPETENCY ASSESSMENT IN FOCUSED CARDIAC ULTRASOUND—CAN THE USE OF SEQUENTIAL TESTING HELP TAILOR TRAINING REQUIREMENTS?

OBJECTIVES: Varying numbers of scans are required by different professional bodies before focused cardiac ultrasound (FCU) competence is assumed. It has been suggested that innovation in the assessment of FCU competence is needed and that competency assessment needs to be more individualized. We report our experience of how the use of sequential testing may help personalize the assessment of FCU competence.

DESIGN: This was a planned exploratory reanalysis of previously prospectively collected data. FCU was performed sequentially by an intensive care trainee and expert on the same patient. Assessment of left ventricular (LV) function by the trainee and expert was compared. Sequential testing methods were used in the analysis of this data to see if they could be used to help in the assessment of competence. Each trainee had completed a 38-hour teaching program and a logbook of 30 scans prior to enrollment.

SETTING: Tertiary Australian not for profit private academic hospital.

MEASUREMENTS AND MAIN RESULTS: Two hundred seventy paired echocardiograms were completed by seven trainees. For trainees to achieve greater than 90% accuracy in correctly assessing LV function when compared with an expert, a variable number of scans were required. This ranged from 13 to 25 (95% CI, 13–25) scans. Over the study period, the ability to correctly identify LV function was maintained, and it appeared there was no degradation in skill.

CONCLUSIONS: Using the Sequential Probability Ratio Test demonstrates a variable number of scans were required to show greater than 90% accuracy in the assessment of LV function. As such, the use of sequential testing could help individualize competency assessments in FCU. Additionally, our data suggests that over a 6-month period, echocardiographic skill is maintained without any formal teaching or feedback. Further work assessing the utility of this method based on larger samples is required.

KEY WORDS: assessment; competency; echocardiography; intensive care; skill retention

FOCUSED CARDIAC ULTRASOUND (FCU) is a mandated skill by some professional bodies, such as the College of Intensive Care Medicine (CICM) in Australia and New Zealand. As such, it forms part of the syllabus for these intensive care trainees. FCU aims to answer focused questions that directly assist with assessment and management of the critically unwell patient by providing real-time anatomical and functional information (1–5).

The number of scans and assessment standards required to determine competence in FCU varies widely (6), and the various pathways have been summarized by Flower et al (7). Problems with using a fixed number of scans to determine competence have been highlighted. Brooks et al (8) found significant heterogeneity in the skill levels of trainees following the completion of
30 assessed scans and questioned how best to assess individual competence. The American Society of Echocardiography states that the number of scans performed or interpreted is only a surrogate for competency (4). It has also been suggested that innovation in the assessment of echocardiography competence is needed following the COVID-19 pandemic, highlighting difficulties with “number-/volume-based” approach to documenting echocardiographic skill (9).

Sequential testing was considered as a quantitative method, which could be employed in the realm of competency assessment. This study reports our experience as to whether the use of sequential testing could be used to help individualize the assessment of FCU competence.

Within healthcare, sequential testing has been trialed in the assessment of learning curves and trainee competence in multiple specialties (10–15). This includes critical care (16) and ultrasound training and competence (17–19), including the visual assessment of left ventricular (LV) function (20).

Sequential testing was first reported by Wald (21) in the United States and Barnard (22) in the United Kingdom and known as the “Sequential Probability Ratio Test (SPRT).” A technique based on SPRT, known as “cumulative sum (CUSUM) test,” was published by Page in 1954 (23). SPRT allows the results of a process to be monitored over multiple occurrences. Unlike conventional statistical hypothesis testing, which is performed at the conclusion of a series of trials, SPRT is performed immediately after each observation. This means that no additional tests are required other than those needed to decide to accept or reject the null hypothesis of competence (24). Conversely, CUSUM was designed for continual testing. It terminates when the process is demonstrably out of control (e.g., a sufficient number of errors have been made to flag a lack of competence) (25, 26).

MATERIALS AND METHODS

This is a planned further analysis of prospectively collected data. The data was collected from a single-center, not-for-profit private 26-bedded CICM-accredited academic ICU in Melbourne, Australia. The full methods of data collection have previously been published by Brooks et al (8). Ethics approval for data collection was granted (EH2016-133) by the Epworth HealthCare Human Research and Ethics Committee. Written consent was obtained from both trainees and patients prior to the data being collected.

All patients in the ICU or coronary-care unit who were not expected to be discharged within the next 2 hours were eligible if there were no exclusion criteria met. Exclusion criteria were atrial fibrillation, subcostal or intercostal drains, pneumothorax, or being deemed inappropriate by the treating intensivist.

Independent, blinded scans were carried out by seven intensive care trainees, who had completed a CICM-accredited FCU course and a logbook of 30 scans prior to enrollment. During the study, the ability for a trainee to correctly classify LV function when compared with an expert was assessed. An expert was either a cardiac sonographer with greater than 5-years clinical experience or an ICU consultant with a Diploma in Diagnostic Ultrasound. The second scan was performed immediately after the first, and the trainee was blinded to the expert’s assessment. The trainees received no instruction or assistance with any part of the scan and did not receive any feedback during data collection. The scans were carried out over a 5-month period, and 270 paired echocardiograms were performed. The pro forma for data collection is included as Appendix 1 (http://links.lww.com/CCX/B13).

Each trainee’s performance was examined separately using their obtained error pattern. An error was recorded if there was a discrepancy between the trainee and expert’s recording of LV function. This was dichotomized into normal and mild LV impairment versus moderate and severe impairment.

There are four key values that are required for assessing a process using SPRT (24), and the values used within this study are as follows:

1) Alpha, α, was set at 0.05 (the probability of rejecting the null hypothesis if it is in fact true).
2) Beta, β, was set at 0.1 (the probability of accepting or failing to reject the null hypothesis if it is in fact false).
3) p1 was set at 0.1. This is defined as that probability of error or misclassification below which would be defined as acceptable (an error rate of less than 10% in classifying LV function when compared with an expert was regarded as acceptable).
4) p2 was set at 0.25. This is defined as the probability of error above which would be defined as unacceptable (an error rate >25% in classifying LV function when compared with an expert was regarded as unacceptable).

These four values are used to define h1 and h2, which act as boundaries for acceptable and unacceptable performance.
There are three options, based on the outcome of each scan:
1) Greater than 90% accuracy (<10% error), in which case stop testing (no further scans)
2) Less than 75% accuracy (>25% error), in which case stop testing (no further scans)
3) No decision can be made (error rate is 10–25%) so at least one further scan is required

In addition to the SPRT procedure described above, CUSUM was employed to monitor the error rate over the entire sequence of scans to ascertain if the error rate later increased, even after competence had initially been shown using SPRT. Once a trainee exhibited satisfactory performance (based on SPRT), the remaining trials were monitored using CUSUM to ascertain whether the error rate became unacceptable subsequently. CUSUM was employed with the same values as SPRT (12).

All statistical analyses were performed using Stata 16 (Stata Corporation, College Station, TX, 2019), with SPRT being programmed by biostatistician (D.P.M.), within Stata; 95% CIs for proportion of errors were obtained using the Clopper-Pearson method (27).

RESULTS

Table 1 shows the patient characteristics.

The use of SPRT showed that in order for a trainee to achieve greater than 90% accuracy when assessing LV function compared with an expert, a variable number of scans were required. This is shown in Table 2. Greater than 90% accuracy occurs when the “accept” boundary is crossed on SPRT testing. This ranges from 13 to 25 scans, with a median of 13 scans (interquartile range, 13–25 scans). Individual SPRT charts are shown in Figure 1.

If a fixed number of scans were required for a trainee to demonstrate greater than 90% accuracy in the assessment of LV function, a minimum of 25 scans would be suggested based on the upper limit of the 95% CI for the median number of scans (13, 95% CI, 13–25 scans) for the seven trainees.

There appears to be no pattern as to when errors were observed. Thus, it appears that, over the study period, there is no degradation in skill, and the ability to correctly identify LV function is maintained. Additionally, the results of the CUSUM analyses suggested that none of the trainees exhibited unacceptable performance at any stage, even after acceptable performance had been established.

### Table 1. Patient and Scan Characteristics

| Characteristic                        | n (%)   |
|---------------------------------------|---------|
| Age, yr (SD)                          | 66.7 (12.2) |
| Male                                  | 158 (64.2) |
| Location                              |         |
| ICU                                   | 120 (44.0) |
| Coronary Care Unit                    | 150 (56.0) |
| Intubated                             | 16 (6.0) |
| Admitting unit                        |         |
| Cardiology                            | 106 (43.1) |
| Cardiothoracic                         | 56 (22.8) |
| Neurosurgery                          | 18 (7.3) |
| Orthopedics                           | 16 (6.5) |
| General medicine                      | 12 (4.9) |
| Other                                 | 38 (15.4) |
| Mean scan time, min (95% CI)*         | 25.3 (22.5–28.1) |
| Focused cardiac ultrasound findings (on expert scan) | 68/269 (25.3) |
| Abnormal left ventricular function     | 17/261 (6.5) |
| Presence of pericardial effusion      |         |

*Adjusted for clustering within trainees.

DISCUSSION

Our study demonstrates sequential analysis could help in personalizing the training experience of those trainees who have recently learnt FCU. When trainee scan performance was analyzed using SPRT, a type of sequential analysis, variability in the number of scans required to determine greater than 90% accuracy (from 13 to 25 scans) in the correct assessment of LV function was demonstrated. As previously discussed, current competency assessments in FCU require the completion of a fixed number of scans prior to competency being assumed. The variability demonstrated in this study (even in a group of trainees who had completed 30 scans prior) is important because it is possible that the use of sequential analysis could be used to tailor the training requirements of a trainee. For example, their use may enable a trainee to sit a summative competency assessment at an appropriate time in their training when they demonstrate accurate FCU assessment when compared with an expert. This could improve efficiency in FCU training and assessment, which is increasingly important with an increasing number of
trainees and associated teaching requirements. The use of a summative assessment provides a safety net by ensuring that all facets of the FCU examination are performed to an acceptable standard as opposed to just those facets being assessed using sequential analysis. This would help alleviate some of the criticism of using sequential tests in medical assessment, namely it might lead to “false reassurance” because early successful performance may “mask subsequent ineptitude” (28). The use of sequential analyses in this manner also allows for a quantitative method to suggest further education may be required if trainees are meeting the lower accuracy boundary, rather than just suggesting the “easier” recommendation of more studies (9). This again may help streamline valuable teaching resources to those who need it most.

Defining competence is the key to credentialling programs worldwide. Without being certain of what is required to determine competency, those charged with credentialling are left with difficult decisions. There appear to be no other reported uses of sequential testing being employed in FCU competence assessment in this manner. One study (20) in which novice practitioners watched cine loops of varying LV function demonstrated approximately 50 cases with feedback were required to become “proficient” at eyeballing ejection fraction. Again, using sequential analysis, variability in the numbers of cine loops that each novice practitioner required was shown, ranging from 40 to 71 cases. Our study differs, however, as trainees had to obtain an appropriate FCU image on a critical care patient and subsequently interpret it in real time, as opposed to being shown prerecorded cine loops. This is important to consider as FCU competency requires image acquisition, in addition to analysis and interpretation (4). It has been shown differing number of studies are required to be considered competent with image acquisition compared with competence in interpretation and integration (29).

Sequential testing could subsequently be employed to ensure ongoing competence. Skill retention is an important part of maintaining competence. Evidence surrounding skill retention of FCU is limited (30). Knowledge retention and image interpretation at 1 year have been demonstrated (31, 32); however, when scanning ability is assessed, it has been observed that degradation occurs as soon as 1 month (33). Knowledge of this is an important factor in order to adequately assess maintenance of FCU competence. CUSUM analyses suggest that, in this cohort of trainees, no one exhibited unacceptable performance. If someone who has previously been deemed competent then demonstrates unacceptable performance through the use of sequential testing, then this could form part of a revalidation assessment and could potentially trigger a requirement for additional training.

This method does have limitations. This data is from a single center that has been retrospectively reanalyzed, and the numbers of trainees involved are small. There was also a low proportion of intubated patients (6%). A criticism of SPRT and CUSUM is that depending on the set values for α, β, p1, and p2, then the number

| Trainee | Number of Scans Required to Demonstrate >90% Accuracy Assessing Left Ventricular Function Using Sequential Probability Ratio Tests | Cumulative Number of Errors at This Scan Number | Scan Number at Which First Error Observed | Total Scans | Total Correct | % Correct (95% CI) |
|---------|-------------------------------------------------------------------------------------------------|-----------------------------------------------|------------------------------------------|-------------|---------------|-----------------|
| 1       | 13                                                                                             | 0                                             | 15                                       | 41          | 38            | 92.7 (80.1–98.5) |
| 2       | 25                                                                                             | 2                                             | 2                                        | 35          | 33            | 94.3 (80.8–99.3) |
| 3       | 19                                                                                             | 1                                             | 7                                        | 34          | 33            | 97.1 (84.7–99.9) |
| 4       | 25                                                                                             | 2                                             | 6                                        | 36          | 33            | 91.7 (77.5–98.2) |
| 5       | 13                                                                                             | 0                                             | 14                                       | 44          | 42            | 95.4 (84.5–99.4) |
| 6       | 13                                                                                             | 0                                             | 26                                       | 40          | 39            | 97.5 (86.8–99.9) |
| 7       | 13                                                                                             | 0                                             | 31                                       | 35          | 34            | 97.1 (85.1–99.9) |
Figure 1. Sequential Probability Ratio Test charts for each trainee. Each point represents a trainee’s focused cardiac ultrasound (FCU) assessment of left ventricular (LV) function compared with an expert. The lower “accept” line depicts an accuracy greater than 90% when compared with an expert echocardiographer. Each FCU that does not agree with the expert’s assessment of LV function (“an error”) moves +1 on the y-axis. If too many errors are made resulting in an accuracy of less than 75%, the upper “reject” line is crossed. If the accuracy is between 75% and 90%, no decision can be made, therefore, further FCUs are required. This is the area between the upper and lower boundaries.
of scans needed to demonstrate > 90% accuracy can vary. The smaller the difference between p1 and p2, the larger the number of scans required. Using a value of 0.1 for p1 is consistent with reported interobserver variability when assessing LV function (34). This would need to be considered prior to wider application. The choice of which test is employed is also important. A comparison of resetting-SPRT (a more complex form) and Learning Curve-CUSUM observed that the two procedures obtained different results (35). Further investigation is required into the practicality and limitations of both simple (as used in this study) and more complex sequential testing techniques applied in this setting.

Within this study, sequential testing has been used in the assessment of accurately identifying LV function only. Correct assessment of LV function has previously been shown to be accurately assessed by novices (8, 32). We do not report on using this method to assess other parts of the FCU examination whose agreement between expert and trainee has been shown not to be as good, for example, inferior vena cava diameter (8). The optimal number of parts of the FCU examination a trainee would need to be deemed accurate on prior to sitting any assessments would need to be established.

This study does not address any of the logistical issues surrounding the use of sequential analyses in this manner. There would need to be a large amount of further work prior to this being used in this manner. For example, discussion around what “expert standards” are appropriate for a trainee to compare their scans with would need to occur. Additionally, an easily accessible method of logging scan performance and conducting sequential analysis would need to be created. This could be in the form of a mobile application (36). Finally, this study does not report on the education impact following the application of sequential analysis, and further work would need to assess how or if the use of sequential analyses changes the certification and credentialling process.

**CONCLUSIONS**

In this single-center trial involving seven trainees, a variable number of scans to demonstrate 90% accuracy in the assessment of LV function was shown when SPRT was used. Given this variability, it is possible that the use of sequential testing could help individualize competency assessments in FCU. Additionally, it also appears that over a 6-month period, echocardiographic skill is maintained without any formal teaching or feedback.

Further prospective work, with larger samples, is needed to compare and refine different methods of sequential testing in this context and to see how it could be integrated reliably into assessing FCU competence. A prospective trial comparing sequential analysis with an approach using a fixed number of scans in novice FCU trainees could be the logical next step.

**ACKNOWLEDGMENT**

We thank the ICU sonography research staff Katrina Timmins and Karen Scholz.

1. Intensive Care Unit, Epworth HealthCare, Melbourne, VIC, Australia.
2. Research Development and Governance, Epworth HealthCare, Melbourne, VIC, Australia.
3. Department of Health Sciences and Biostatistics, Swinburne University of Technology, Melbourne, VIC, Australia.
4. Department of Medical Education, University of Melbourne, Melbourne, VIC, Australia.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s website (http://journals.lww.com/ccxjournal).

Supported, in part, by a grant from the Epworth Research Institute and registered (Trial registration: NCT02961439) for data collection.

The authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: hgmwalker@me.com

This work was performed at Epworth Richmond, 89 Bridge Rd, Richmond, VIC 3121, Australia.

**REFERENCES**

1. McLean AS: Echocardiography in shock management. Crit Care 2016; 20:275
2. Vignon P: What is new in critical care echocardiography? Crit Care 2018; 22:40
3. Vieillard-Baron A, Millington SJ, Sanfillipo F, et al: A decade of progress in critical care echocardiography: A narrative review. Intensive Care Med 2019; 45:770–788
4. Spencer KT, Kimura BJ, Korcarz CE, et al: Focused cardiac ultrasound: Recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr 2013; 26:567–581
5. Nanjayya VB, Orde S, Hilton A, et al: Levels of training in critical care echocardiography in adults. Recommendations from
