Assessment of the clinical utility of an ultrasonic monitor of cardiac output (the USCOM) and agreement with thermodilution measurement

Rob A Phillips

To the Editor: Is it strange that Boyle et al[1] cannot get continuous-wave Doppler, and more specifically USCOM (ultrasonic cardiac output monitor) (USCOM Ltd, Sydney, NSW), to agree with circulatory measurements by pulmonary artery catheter (PAC)?

Continuous-wave Doppler, the USCOM method, is a routine part of the echocardiographic examination, and has been in clinical use for over 40 years as a gold-standard measure of blood flow. The American Heart Association, American College of Cardiology and American Society of Echocardiography, in their combined guidelines, endorse Doppler as a measurement method with high levels of clinical proof and conclude:

> The ability of Doppler echocardiography to provide unique noninvasive information with minimal discomfort or risk without using contrast material or ionizing radiation, coupled with its portability, immediate availability, and repeatability, accounts for its use in virtually all categories of cardiovascular disease.[2]

Specifically, USCOM has been validated with electromagnetic flow probes, PAC, Fick's method, echocardiography, artificial hearts and external cardiac pumps, inert gas rebreathing and magnetic resonance imaging in patients aged from 26 weeks' gestational age to 87 years, and from 0.12 L/min to 18.7 L/min. Further, 31 of all 37 USCOM validation studies (91%) concluded positively, while 6 of 8 studies cited by Boyle et al (75%) also had positive conclusions. Perhaps the final word in USCOM validation belongs to Jain et al[3] from the Cedars Sinai ICU, the home of Swan and Ganz and their catheter. From their comparison of PAC and USCOM, similar to that of Boyle et al, they concluded: "[USCOM's] noninvasiveness and short learning curve suggest that this device is a good substitute over the PAC for determining cardiac function. in the [surgical] ICU". This decisive paper from the "home" of the PAC contradicting their results was not cited by Boyle et al.[1]

Recent evidence from Washington University[4] identifies USCOM as having a 91% positive predictive value for identifying fluid responsiveness in patients receiving ventilation, in atrial fibrillation and on vasopressor therapy, suggesting it is nearly twice as predictive as invasive central venous pressure (CVP) and mean arterial pressure (MAP) and can be used in most patients. In an Australian study of patients with sepsis, USCOM identified fluid responders more than twice as effectively as B-type natriuretic peptide level or CVP.[5] This clinical effectiveness has not been achieved by PAC since its introduction in 1971.[6] Surely, given the challenges of ICU fluid management, doubling the identification of fluid responsiveness non-invasively is clinically useful.

Sixteen per cent of cardiac surgical patients in the study by Boyle et al had aortic valve replacements, thereby ensuring disagreement between the two methods. USCOM specifically recommends an alternative approach for monitoring patients with aortic valve replacements — an approach not adopted in this study. Tricuspid regurgitation is a proven source of PAC error, and will be present with mild plus grading in most cardiac surgical patients, thus compromising the PAC–USCOM correlation. No assessment of the grade of tricuspid regurgitation was made in these patients, nor was it mentioned as a limitation of the study. Further, the possibility that PAC might be inaccurate was not raised by the authors, despite multiple studies citing PAC inaccuracy, and 40 years of clinical use without evidence of outcome benefit.[6]

No comparative mortality or morbidity data were presented by Boyle et al, either from their study or from other studies. Yet such evidence is critical to the question of clinical applicability and while PAC is associated with a constellation of common and serious complications, USCOM has none.

The effectiveness of USCOM, like most technologies, is a function of education and experience. There are challenges for new USCOM users who are unaccustomed to basic echocardiography practices, as they develop the manual skills required to acquire Doppler signals. Boyle et al reported a 28% success rate for acquiring acceptable transthoracic Doppler signals,[1] a result totally inconsistent with prior echocardiographic and USCOM studies, and an indicator of the level of competence of the operator and a predictor of the outcome of the study. By comparison, a review of 30 studies reporting USCOM feasibility in 1305 predominantly post-cardiac surgical patients (as were those of Boyle et al), demonstrated a mean 93% feasibility for acquiring acceptable signals. Further, USCOM is currently used routinely by a range of specialist physicians, nurses and

technicians in Australia, Asia, Europe and the United States to appropriately and reliably guide clinical care non-invasively, suggesting it is both reliable and feasible.

Considering the wealth of uncited positive evidence for USCOM, and the evolving global adoption of USCOM and Doppler ultrasound, as well as the abundant and persistent evidence of associated risk and ineffectiveness of PAC (also uncited by Boyle et al), it was a courageous conclusion to state that USCOM was an unsuitable tool for use in the ICU compared with PAC! A more scientific conclusion from this study is that, in their hands and in their patients, there was disagreement between the results obtained by a single poorly skilled operator of USCOM and those from a (presumably) experienced PAC operator. If the authors wished to then speculate as to which method wasn't suitable for ICU use, then based on this experiment and current clinical evidence this would have to be the PAC. Presumably, had the study involved echocardiography using an unskilled echocardiographer, and using the current logic, Boyle et al would have concluded that there was no place for echocardiography in the ICU!

Finally, there was no conflict of interest statement in the article by Boyle et al,[1] despite at least one author having an association with a competitive monitoring product. This may have been an oversight by the authors or an editorial omission, but should elicit caution when evaluating a study that reaches such an assertive conclusion.

Editor's note: this letter is a modification of a previous letter to which the response by Boyle and colleagues (below) was originally written.

Competing interests

Rob Phillips is an executive of and shareholder in USCOM Ltd.

Rob A Phillips, PhD Candidate,[1] and Director of Clinical Science and Executive Chairman[2]

1 School of Medicine, University of Queensland, Brisbane, QLD.

2 Department of Clinical Science, USCOM Ltd, Sydney, NSW.

rap.echo@bigpond.com

References

1. Boyle M, Steel L, Flynn GM, et al. Assessment of the clinical utility of an ultrasonic monitor of cardiac output (the USCOM) and agreement with thermodilution measurement. *Crit Care Resusc* 2009; 11: 198-203.
2. Cheitlin MD, Alpert JS, Armstrong WF, et al. ACC/AHA guideline for the clinical application of echocardiography. *Circulation* 1997; 95: 1686-744.
3. Jain S, Allins A, Salim A, et al. Non-invasive Doppler ultrasonography for assessing cardiac function: can it replace the Swan–Ganz catheter? *Am J Surg* 2008; 196: 961-8.
4. Thiel SW, Kollef MH, Isakow W. Non-invasive stroke volume measurement and passive leg raising predict volume responsiveness in medical ICU patients: an observational cohort study. *Crit Care* 2009; 13: R111.
5. Sturgess DJ, Pascoe RLS, Scalia G, Venkatesh B. A comparison of transcutaneous Doppler corrected flow time, b-type natriuretic peptide and central venous pressure as predictors of fluid responsiveness in septic shock: a preliminary evaluation. *Anaesth Intensive Care* 2010; 38: 336-41.
6. Shah MR, Hasselblad V, Stevenson LW, et al. Impacts of the pulmonary artery catheter in critically ill patients: meta-analysis of randomized trials. *JAMA* 2005; 294: 1664-70. ❑

Brendan E Smith and Veronica M Madigan

To the Editor: We feel we must take issue with Boyle et al[1] in their conclusions about the role of continuous-wave Doppler (CWD) ultrasound in the form of the USCOM (ultrasonic cardiac output monitor) (USCOM Ltd, Sydney, NSW) in intensive care units.

Comparing methods for measuring cardiac output is complex, and the pitfalls of bias and precision as a tool for comparison are highlighted by this study. Boyle et al incorrectly assume that a highly variable reference method — in this case, thermodilution (TD) measured by pulmonary artery catheter or by a thermister-tipped femoral or axillary artery catheter (Pulsiocath, Pulsion Medical Systems, Sydney, NSW) — can be used in the Bland–Altman method to compare TD with CWD in the form of the USCOM. In fact, the usefulness of the Bland–Altman comparison is limited by the use of a reference standard with high variability.

Several investigators have shown that the variability for TD measures is inherently high. This is usually addressed in clinical practice by taking an average of three to five readings. However, Boyle and colleagues state that "Three good quality thermodilution curves were averaged to obtain TD cardiac output". This implies that several more curves were produced, and only three that showed optimal morphology and good agreement were selected. While this is a common practice clinically, it may be inappropriate in the research situation, where the true variability should be used, with perhaps only true statistically defined outliers being rejected. It might be argued that the same method should be applied to the contrasting method of measurement, with an equal number of CWD samples of 6–10 heart beats' duration being taken and the "best" three being averaged. But we believe this to be scientifically poor practice, as it represents inherent bias in measurement selection.

Intraclass correlation is intended for paired measures, yet the measures are not comparable. Both TD and CWD could measure cardiac output totally accurately over their respective time frames, and yet show poor correlation. This is analogous to standing beside a motorway and comparing the average number of vehicles passing in three (from perhaps six) pooled

10–20-second observation periods spread over 5–10 minutes with the result of the count from a single 2–3-second "snapshot". Using this method, it is improbable that even heart rate would show good agreement.

Bland and Altman[2] noted this limitation themselves when describing their method:

> Repeatability is relevant to the study of method comparison because the repeatabilities of the two methods of measurement limit the amount of agreement which is possible. If one method has poor repeatability — i.e. there is considerable variation in repeated measurements on the same subject — the agreement between the two methods is bound to be poor too. When the old method is the more variable one, even a new method which is perfect will not agree with it.

With such a highly variable reference measure as TD, even a perfect comparison method would be highly unlikely to achieve less than the 30% error in agreement that the authors considered acceptable.

Importantly, this variability represents both the measurement variability of the method (which is what we are trying to compare) and the physiological variability of cardiac output in the study participants. Differentiating high physiological variability and low measurement variability from low physiological variability and high measurement variability is exploited in studies in which normal participants are studied at rest. The inaccuracy of TD in the presence of tricuspid regurgitation, for example, is well described as a limitation of the TD method, and yet the study by Boyle et al would have included a large but unspecified number of patients with tricuspid regurgitation. After cardiac surgery, patients would also tend to have high physiological variability, particularly those with atrial fibrillation or with intra-aortic balloon pumps in situ (23% in each case). Using a method with high measurement variability in such physiologically variable participants would mean that achieving 30% precision would be almost statistically impossible! With regard to general ICU patients in the study by Boyle et al, we are told only that 45% were "septic" (presumably meaning they had septic shock), 35% were "medical" and 20% were "surgical". The probability of high physiological variability in this group is clearly also considerable.

Furthermore, while we accept that post-cardiac surgery and general ICU patients are difficult CWD subjects, it is striking that only 28% of the Doppler profiles were assessed by an experienced observer as being of even acceptable quality. This represents a failure rate of over 70% in obtaining Doppler data of diagnostic quality. The reason for this is apparent when we learn that the Doppler studies were performed by one observer with experience of "65 ultrasonic cardiac output assessments before the study". The authors do not state how many of these assessments had been performed with the USCOM monitor used in the study or how many of these assessments had been vetted as being of acceptable standard. Experienced echocardiography-trained clinicians would regard 65 studies as extremely inadequate in this class of patient for basic clinical use, let alone for research purposes, where we have a duty of care to our patients both current and future to exercise due scientific rigour.

Brendan E Smith, Associate Professor[1,2]
Veronica M Madigan, Senior Lecturer[1]
1 School of Biomedical Sciences, Charles Sturt University, Bathurst, NSW.
2 Department of Anaesthesia and Intensive Care, Bathurst Base Hospital, Bathurst, NSW.
brendanprivate@hotmail.com

References

1. Boyle M, Steel L, Flynn GM, et al. Assessment of the clinical utility of an ultrasonic monitor of cardiac output (the USCOM) and agreement with thermodilution measurement. *Crit Care Resusc* 2009; 11: 198-203.
2. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10. ❑

Martin Boyle, Liz Steel, Gordon M Flynn, Margherita Murgo, Lisa Nicholson, Maureen O'Brien and David Bihari

In reply: Thank you for the opportunity to respond to the letters from Phillips* and from Smith and Madigan.

The purpose of our study[1] was to assess the clinical utility of the USCOM (ultrasonic cardiac output monitor) (USCOM Ltd, Sydney, NSW) in clinical practice in critical care patients. Accordingly, we assessed the reliability of USCOM in tracking cardiac output (CO) compared with the clinical standard method of bolus thermodilution (TD) with regard to its ability to identify low, normal and high CO and to track changes. We also documented the rate of failure to obtain an USCOM waveform, and then further subjected the waveforms obtained clinically to an ideal standard as assessed by an independent assessor.

The validation of the USCOM using flow probes, referred to by Phillips, was an animal study by Critchley et al[2] in which dog aorta diameter, measured at thoracotomy, was used in USCOM calculation of CO. In humans, aortic valve diameter is calculated from a formula based on work by Nidorf et al.[3] Critchley et al discussed this calculation in their article, reporting that a 95% confidence interval on the nomogram

used in humans represents a ±10%–20% variation in aortic diameters, and pointing out that even a small discrepancy in diameter can cause quite large systematic errors (calculation of valve area πr^2) in CO measurements. This issue had also been raised in a recent clinical study in humans by Van den Oever et al,[4] who stated that, based on the confidence limits of the estimated valve area, the accuracy of CO measurements could not be expected to exceed ±40%. Van den Oever and colleagues were unable to show agreement between CO measured by the USCOM and by pulmonary artery TD with an a priori bias of less than 0.5 L/min and a limit of agreement of ±1 L/min. Their study was not referred to by Phillips. Our results showing poor agreement between USCOM and TD were also consistent with another recent study in 89 cardiac surgical and intensive care patients.[5]

While the accuracy of measurement is no doubt operator-dependent (and also influenced by patient factors), it is very much an overstatement to blame the observed lack of agreement between the two measurement methods on the operator when the machine systematic error is as described above. The error in the estimation of valve area (and thus CO) applies to measurement between individuals. Measurements in the same patient will not be affected by this error, although operator- and patient-dependent factors may still be an issue, and, as pointed out by Phillips, the USCOM has been used to assess volume responsiveness.[6]

Phillips asks why we did not refer to the study by Jain et al,[7] which he described as "decisive". This study, published in December 2008, was not available to us at the time of writing up our study.[1] The study by Jain et al, involving 31 adult surgical intensive care unit patients, had a total of 120 paired measurements. It had a blinded and an unblinded phase. In the blinded phase, the percentage limit of agreement was 22.8%, using an a priori standard of less than 30% (as we did in our study). However, there were 60 paired observations in only 14 patients in this phase. Multiple comparisons in the same individual have the potential to significantly reduce the measurement error of the USCOM, and therefore the limits of agreement would be expected to be narrower compared with a study like ours, in which each individual was studied only twice. The study by Jain et al[7] may also be limited in the assessment of any failure rate to obtain Doppler waveforms as, presumably, multiple measures were made in patients where acceptable waveforms could be obtained.

Phillips raised concerns about the number of patients who had undergone valve repair. In patients who had an aortic valve replacement, the pulmonary valve was used for USCOM measurements. Cardiac surgical patients with documented moderate or severe tricuspid regurgitation were excluded from the study. Tricuspid valve function was not specifically assessed in the intensive care patients.

Our article reported a 16% failure rate in obtaining a Doppler waveform. This is a rate consistent with other reports and with information provided by USCOM Ltd (http://www.uscom.com.au/support/faqs.html [accessed Apr 2010]). In our study, the waveforms were further independently assessed against a high standard. Another interpretation of these results is that only 28% of waveforms that could reasonably be obtained in clinical practice in this patient group met an ideal standard.

On the question of the competence of the USCOM operator in our study, we stand by the statement in our article that the operator had received standardised training. The training was conducted by an USCOM representative, reinforced by clinical experience, and further followed up by an USCOM representative. The initial clinical practice of the USCOM operator in our study was part of an assessment of ease-of-use and interoperator reproducibility presented at a conference, the abstract of which is available on the USCOM website (http://www.uscom.com.au/benefits/evidence.php?cat=26 [accessed Apr 2010]). As stated in our article, the USCOM operator had undertaken at least 65 USCOM measurements before the study.

With regard to a conflict of interest statement, the following was provided on submission of our article for publication: "Potential conflicts of interest with regard to the submitted manuscript are:

- USCOM Ltd Sydney provided an USCOM device for the study;
- Associate Professor David Bihari was, during the period of data collection, a non-executive director of Pulsion Pacific Ltd."

Smith and Madigan raise valid concerns about the difficulties of comparing measurement methods where differing time windows, measurement variability and background biological variability are confounding issues. Our study used well accepted methods in assessing the two measurement techniques. Bolus TD is a clinical standard, rather than a gold standard, with well understood limitations. Taking the average of 3–5 good-quality waveforms reduces variability and increases accuracy. This is analogous to rejecting poor-quality waveforms obtained with the USCOM using FlowTracer software or selecting good-quality waveforms using the TouchPoint method.

Bland–Altman analysis is used when there is no gold standard — that is, CO is taken as the average of the values obtained by the methods being assessed. We used this analysis appropriately, and used an a priori standard for agreement. Variability of the USCOM is perhaps the issue. We agree with Smith and Madigan that it is very difficult to demonstrate agreement between measurement methods when one (USCOM) has a systematic error of up to 40%, without considering operator- and patient-dependent

issues. This is very different to the "better than 95%" accuracy stated by Smith in his testimonial on behalf of USCOM Ltd (http://www.uscom.com.au/benefits/test_brendan.html [accessed Apr 2010]).

We appropriately used correlation to assess percentage changes in CO, as these are dimensionless units.

We believe the concerns of Smith and Madigan about the competence of the USCOM operator are addressed above. It is pertinent to note that in response to the question "Is it easy to use?", Smith stated: "My grand daughter is six years old and she can use it. It really is child's play" (http://www.uscom.com.au/benefits/test_brendan.html [accessed Apr 2010]). We maintain that our results and a balanced review of the current peer reviewed literature support the conclusion of our study.

* *Editor's note:* the letter by Phillips that Boyle and colleagues respond to here was an earlier draft of the one published above.

Author details

Martin Boyle, Clinical Nurse Consultant
Liz Steel, Fellow in Intensive Care
Gordon M Flynn, Fellow in Intensive Care
Margherita Murgo, Nurse Educator
Lisa Nicholson, Nurse Educator (Acting)
Maureen O'Brien, Clinical Nurse Educator
David Bihari, Associate Professor and Senior Staff Specialist
Department of Intensive Care, Prince of Wales Hospital, Sydney, NSW.
martin.boyle@sesiahs.health.nsw.gov.au

References

1. Boyle M, Steel L, Flynn GM, et al. Assessment of the clinical utility of an ultrasonic monitor of cardiac output (the USCOM) and agreement with thermodilution measurement. *Crit Care Resusc* 2009; 11: 198-203.
2. Critchley LA, Peng ZY, Fok BS, et al. Testing the reliability of a new ultrasonic cardiac output monitor, the USCOM, by using aortic flowprobes in anesthetized dogs. *Anesth Analg* 2005; 100: 748-53.
3. Nidorf SM, Picard MH, Triulzi MO, et al. New perspectives in the assessment of cardiac chamber dimensions during development and adulthood. *J Am Coll Cardiol* 1992; 19: 983-8.
4. Van Den Oever HLA, Murphy EJ, Christie-Taylor GA. USCOM (ultrasonic cardiac output monitors) lacks agreement with thermodilution cardiac output and transoesophageal echocardiography valve measurements. *Anaesth Intensive Care* 2007; 35: 903-10.
5. Thom O, Taylor DM, Wolfe RE, et al. Comparison of a supra-sternal cardiac output monitor (USCOM) with the pulmonary artery catheter. *Br J Anaesth* 2009; 103: 800-4.
6. Theil SW, Kollef MH, Isakow W. Non-invasive stroke volume measurement and passive leg raising predict volume responsiveness in medical ICU patients: an observational cohort study. *Crit Care* 2009; 13: R111.
7. Jain S, Allins A, Salim A, et al. Noninvasive Doppler ultrasonography for assessing cardiac function: can it replace the Swan–Ganz catheter? *Am J Surg* 2008; 196: 961-8. ❑