Intra- and inter-observer reliability using a noninvasive ultrasound cardiac output monitor in healthy anesthetized children

Sonny Dhanani1,2, Nick J. Barrowman3, Roxanne E. Ward3 & Kimmo T. Murto2,4

1 Department of Pediatric Critical Care, Children’s Hospital of Eastern Ontario, ON, Canada
2 University of Ottawa, Ottawa, ON, Canada
3 Clinical Research Unit, Children’s Hospital of Eastern Ontario Research Institute, ON, Canada
4 Department of Anesthesia, Children’s Hospital of Eastern Ontario, ON, Canada

Abstract

Background: Accurate and reliable evaluation of cardiac index (CI) in critically ill pediatric patients can optimize their management. Although validated, noninvasive ultrasound measurement techniques have been previously shown to be unreliable because of observer variability.

Objective: To confirm intra- and inter-observer reliability when using the noninvasive USCOM® in healthy anesthetized children.

Methods: Prospective observational study at the Children’s Hospital of Eastern Ontario, Ottawa, included newborns to 12 years of age undergoing elective surgery or magnetic resonance imaging. The USCOM® was used to assess CI via aortic flow with a trans-sternal approach. Two trained observers were responsible for taking two measurements of CI each at steady state in randomized succession after stable depth of anesthesia was achieved.

Results: Fifty-nine patients were included. Forty-seven (80%) were between 3 and 7 years old, with 57% male. The mean difference ± sd for repeat CI measurements by each of two observers was 0.11 ± 0.47 and 0.05 ± 0.65 l/min/m², respectively. Intra-observer reliability for these repeat measurements by each observer determined by Lin’s concordance correlation coefficient was 0.92 and 0.85, respectively. The mean difference ± sd between observers was 0.16 ± 0.59 l/min/m², and Lin’s concordance correlation coefficient was 0.87. The two observers subjectively rated measurements as ‘Difficult’ or ‘Very difficult’ only 14% (16/118) and 3% (4/118) of the time, respectively. No adverse events were reported.

Conclusion: This study confirms that the USCOM® is relatively easy to use and reliable in healthy children when operated by trained users.

Introduction

Timely and appropriate assessment of cardiac index (CI) in critically ill pediatric patients can optimize their management for sepsis, trauma, and postsurgical-related cardiovascular compromise (1,2). There are several methods to estimate CI (3–5). Most are difficult, time-consuming, and invasive. Thus, the determination of cardiac index at the bedside is generally considered impractical, especially when a child is unwell. Non-invasive bedside methods to monitor hemodynamics and guide therapy simultaneously have been unreliable.

The USCOM®, Ultrasound Cardiac Output Monitor, (USCOM Ltd, Sydney, NSW, Australia) is a noninvasive, portable, and easy to use monitor that measures CI by assessing aortic valve outflow using...
Doppler ultrasound (Figure 1). The inherent reliability or precision deficiencies (accuracy was not measured here) of this technique are attributed to the inability to attain consistent continuous wave Doppler spectra of ascending aortic flow. Previous studies have validated patient height and weight to valve area in children and neonates (6,7). An automated analysis system, newly validated nomograms, and ease of use are recent improvements to this noninvasive technique.

Noninvasive ultrasound techniques for assessing CI have previously been studied in infants and children in sepsis, trauma, and surgical management. A review by Chew et al. examining the accuracy of ultrasound techniques showed good validity when compared to other invasive techniques such as Fick and thermodilution (8). Several pediatric institutions in the United States, United Kingdom, and Australia are presently undergoing preliminary trials with the USCOM® (9,10). Validity of this tool to measure CI has been well established (7–9). Despite its potential application, routine use by clinicians has been limited because of concerns with reliability (8,11–17).

Reliability or precision is an indication of how well a test or technique measures something in a consistent and reproducible manner in an unchanged population. In particular, intra- and inter-observer reliability describe consistency between measurements made by the same observer and different observers, respectively. The main goals of this study were to confirm intra- and interobserver reliability and ease of use of the USCOM® in healthy anesthetized pediatric patients.

Materials and methods

Subjects
We conducted a prospective observational cohort study of healthy anesthetized children undergoing elective surgery or magnetic resonance imaging (MRI) studies at the Children’s Hospital of Eastern Ontario (CHEO), Ottawa, Ontario, Canada from December 2006 to December 2008. Subjects were eligible if they were newborn to 12 years old and had an ASA (American Society of Anesthesiologists) classification of I or II. To prevent unpredictable deviations in CI, we excluded subjects who had active cardiovascular or pulmonary disease (e.g. active symptoms or history of moderate to severe asthma), were undergoing procedures involving intracranial, thoracic, cardiovascular or upper abdominal surgery, or were currently taking vasoactive medications. This study was approved by the CHEO Research Ethics Board and by Health Canada for Class III investigational testing under the Medical Devices Bureau. Written informed consent was obtained from all study participants.

Prior to patient recruitment, two observers (a pediatric intensivist and a pediatric cardiac anesthesiologist) underwent a training course provided by the manufacturer to gain competency with the equipment and measurement technique. Twenty volunteers each were evaluated as per manufacturer’s recommendation for competency (Fremantle scoring system. (12)). The validated Fremantle score includes six criteria (visual and auditory) suggested for optimal flow tracings. A score of ≥3 out of six is considered an adequate study. None of these subjects were included in the final data analysis.

Consented subjects underwent routine preoperative evaluation and a standardized anesthetic and monitoring technique. All subjects were anesthetized by the same individual. Patients underwent an inhalational induction with 100% oxygen combined with a maximum of 8% inspired sevoflurane. Once intravenous access was achieved, propofol 2 mg·kg⁻¹ was administered and the patient was intubated. Rocuronium 0.6 mg·kg⁻¹ was administered to facilitate intubation at the discretion of the anesthesiologist.

Vital signs were monitored continuously postanesthetic induction. End-tidal carbon dioxide and sevoflurane percent concentration were maintained between 35 and 45 mmHg and 2.9–4.4%, respectively. Oxygen administration was adjusted to maintain saturations >97%. Stable baseline vitals were determined when blood pressure, heart rate, oxygen saturations, end-tidal carbon dioxide levels, and temperature were constant for two successive one-minute interval measurements.

Figure 1 Example of Aortic Flow Tracing on Ultrasound Cardiac Output Monitor (USCOM®).
Patients were not handled during the period of hemodynamic stabilization. Cardiac index measurement data was obtained prior to the scheduled surgical or diagnostic intervention.

To ensure a stable CI throughout the measurement period, blood pressure and heart rate were required to be within 15% of baseline measurements throughout the observation time. If the respiratory or hemodynamic values deviated more than 15% from the baseline measurements, the recording was stopped. A new hemodynamic baseline was established by again measuring constant baseline vitals for two successive one-minute interval measurements. The USCOM® measurements were then repeated.

Measurement procedure

We used the USCOM to noninvasively assess Doppler flow across the aortic valve. Blood flow was measured by using a transducer/probe (3.3 MHz for children weighing <20 kg and 2.2 MHz if >20 kg) placed on the chest in the suprasternal position. Cardiac output (CO) = stroke volume (SV) X heart rate (HR) where

\[
SV = \text{volume time interval (VTI) (cm)} \times \text{cross-sectional area (XSA) (cm}^2) \]

The distance that a column of blood travels with each stroke is calculated using a unique Touchpoint® automatic flow profile trace, which represents the VTI. Cross-sectional area of the valve is derived from height-indexed regression equations. VTI and HR are measured and thus operator dependent while XSA is derived from validated nomograms.

The same two observers (Observer A and B) were responsible for taking two measurements (Time 1 and 2) each in randomized succession determined by a computer generated random allocation chart, for a total of four measurements for each subject. Measurements were taken at either end-expiration or inspiration if the patient was undergoing positive pressure or spontaneous ventilation, respectively. The optimal waveform was recorded and, in addition, two waveforms immediately before and after it were recorded to account for CI variation because of ventilation. The five waveforms were averaged. Each measurement was limited to three minutes. To ensure blinding of results, the observer not actively engaged in measuring CI was absent from the room to prevent visual or auditory bias during Doppler flow measurements by the other observer. All values including CI measurements displayed on the screen were hidden from the observer with a cover to prevent bias. After all four measurements were recorded, the study was terminated and the procedure continued without further study involvement. The measurements were interpreted post hoc.

The CI measurements were recorded by a research assistant while the observers were blinded to these results. Standard patient demographic information was collected including diagnosis, reason for surgery, and medications. Ease of assessment and the Fremantle score for each data set was recorded (12). Any adverse events were recorded and reported according to Canadian regulations (18).

Statistical analysis

We determined the sample size using the Bland formula for repeatability and to estimate the study reliability coefficient, assuming a predetermined within-subject standard deviation. The following assumptions were made: (i) a within-subject standard deviation of 10%; (ii) 4 measurements per subject; (iii) 2 observers. The sample size required to estimate the width of the 95% confidence interval within 10% was 64 subjects.

Descriptive summaries of baseline characteristics were generated for all participants. Dichotomous variables were summarized using percentages, normally distributed continuous variables were summarized using means together with standard deviations, and continuous variables that were not normally distributed were summarized using medians together with range.

Coefficients of variation, defined as the standard deviation divided by the mean of the observations, were computed to allow comparison with other reports on the repeatability and reproducibility of cardiac output measurements. The coefficients of variation along with their 95% confidence intervals were generated for both intra- and inter-observers’ measures of the participants’ cardiac index. In addition, reliability (inter- and intra-observer) was assessed using two different statistical techniques: (i) Lin’s concordance correlation coefficient was calculated along with its 95% confidence interval (19). (ii) A plot of the difference of the results for each subject against their mean was generated to detect any possible relationship between the measurement error and the true value. Limits of agreement were calculated using the Bland and Altman method (20). We expected 95% of the differences to be within two standard deviations of the difference, which would be consistent with the definition of a repeatability coefficient adopted by the British Standards Institution (21).

Results

Sixty-nine patients were enrolled between December 8, 2006 and December 22, 2008. Ten patients were subsequently excluded, resulting in 59 patients analyzed.
This was below our desired sample size of 64. One patient was excluded because of incorrect height and weight parameters entered, five were excluded because vital signs were discovered post hoc to have deviated from baseline by more than 15%, two were excluded because one of the observers was unavailable, one was excluded because of cancelation of surgery, and one was excluded because of administering eye drops that were potentially systemically vasoactive.

Table 1 shows the demographic and clinical characteristics of the 59 included patients. Forty-seven patients (79.7%) were between the ages of 3 and 7, 56 (94.9%) had ASA classification I, and 53 (89.8%) had a preoperative diagnosis of dental caries. The cardiac index varied between 3.0 and 8.4 with a mean of 5.5 (SD = 1.2), which is within the realm of normal values for children under sevoflurane induction (4,22).

Table 2 shows the differences in measured CI by observer A and B using the USCOM®.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>sD</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-Observer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference of measurement 1–measurement 2 for CI Observer A</td>
<td>0.11</td>
<td>0.47</td>
<td>0.92</td>
</tr>
<tr>
<td>Difference of Measurement 1–Measurement 2 for CI Observer B</td>
<td>0.05</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>Inter-Observer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference of Observer A12–B12 of CI</td>
<td>0.16</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>Difference of Observer A12–B12 of CI</td>
<td>-0.14</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>Difference of Observer A12 – B12 of CI</td>
<td>-0.20</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>Ease of measurement ≥4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI = Cardiac Index; CCC = Lin’s concordance correlation coefficient.

0.22 l min⁻¹ m⁻²), respectively. Scatterplots of repeat measurements were used for a graphical assessment of reliability. Figure 3 shows the scatterplot for Observer A. The scatterplot for Observer B was similar. Reliability as determined by Lin’s concordance correlation coefficient was 0.92 (95% CI 0.87–0.95) for Observer A and 0.85 (95% CI 0.77–0.91) for Observer B. The coefficients of variation were 8.5% and 11.9% for Observers A and B, respectively.

Interobserver reliability

Interobserver reliability (between two observers) was also assessed. The agreement was similar to the intraobserver comparisons with a mean difference of 0.16 l min⁻¹ m⁻² (95% CI 0.01–0.32 l min⁻¹ m⁻²) between the two observers (Table 2). There was no significant correlation between the interobserver differences and age of the children (P = 0.16). Figure 4 shows the scatterplot for Observer A vs B. Lin’s concordance correlation coefficient was 0.87 (95% CI 0.79–0.92) and the coefficient of variation was 10.7%.

Ease of use

Ease of measurement was assessed for each observation by the nonvalidated Ease of Measurement five-point Likert scale and the validated Fremantle score. Fourteen percent of measurements by Observer A and only 3% of measurements by Observer B were subjectively found to be ‘difficult’ or ‘very difficult’. ‘Difficult’ was subjectively determined to be measurements that required more time, probe manipulation, or extra positioning to obtain adequate Fremantle scores. The rest were deemed average, easy or very easy. The
Fremantle score was $\geq 4/6$ representing adequate visual and auditory tracings in 91% of measurements. Measurements judged by a Fremantle score $\leq 3$ or a subjective estimate of ‘difficult’ or ‘very difficult’ showed slightly lower interobserver agreement.

The mean time for each measurement was 1.10 (SD = 0.34) minutes. There were no noted adverse events.

Discussion

The USCOM$^\text{®}$ uses a patented approach including algorithms directed at determining the flow across the aortic valve. Our study showed that the USCOM$^\text{®}$ was reliable and easy to use in healthy children when operated by trained users. Although we allocated 3 minutes for each measurement, we found measurements only took approximately 1 min each. Ease of use was suggested by very few measurements being subjectively deemed as difficult by the observers. Any difficulties in measurements were not clinically relevant. We showed tight reliability within the same observer and between two observers. (Figures 2–4) The level of variance was low and irrelevant in the clinical setting.

An accurate, reliable, noninvasive method of measuring CI in critically ill children is important. It allows for both monitoring and guiding therapy of hemodynamic parameters in real time. Despite its potential application, its use has been limited because of concerns about reliability even though its validity has been demonstrated when compared to cardiac catheterization and thermodilution assessment of CI.
(23–25). Chew et al. reviewed previous studies that specifically assessed precision or repeatability using other noninvasive ultrasound techniques. These twelve studies assessed 344 children and found up to a 22% difference between measurements by the same observer (intra-observer) and between measurements of two different observers (interobserver) (8). Other studies have shown similar results. These ranges were quite wide and thought to be related to problems with earlier machine estimation of valve size and observer techniques (4,16,17).

One previous study measuring USCOM reliability demonstrated that it was easier to learn with potential for better reliability even in children (26). However, in their study setting, intrinsic variability in cardiac output was a likely confounder. Our study only included children undergoing a standardized anesthetic in an attempt to ensure a constant CI and used blinded randomized methods to assess intra- and inter-observer reliability. Our improved reliability could reflect the attempt to achieve stable hemodynamics, having formally trained operators, and relative ease of use. We separately analyzed the 5 cases that were excluded for vital signs falling outside our 15% deviation parameters. The interobserver differences for all 5 cases were within the previously estimated 95% limits of agreement, suggesting that CI measurements may be reliable during hemodynamic instability.

Limitations to our study include the small number of patients to assess reliability. A sample size of 64 patients was determined after initial findings of a pilot study; however, as a result of post hoc exclusions, we analyzed 59 patients. Larger numbers would have resulted in narrower confidence intervals for the coefficients of variability. It can be argued whether we truly had achieved stable hemodynamics during the CI measurements. We assumed that maintaining the heart rate and blood pressure within 15% of baseline and ensuring stable anesthetic depth, oxygenation, and ventilation in an undisturbed patient could infer stable cardiac output. In addition, we were cognizant of where in the respiratory cycle the measurements were taken and the total measurement time for both observers was short in duration, only four minutes. Ideally, a ‘gold standard’ to measure the CI, such as the Fick method, would have been preferred in the protocol to simultaneously ensure stable CI. This would have complicated the study by requiring cardiac catheterization or other invasive methods. Transthoracic echocardiography is not considered a ‘gold standard’ and likely would not have been a suitable surrogate.

Our study specifically chose healthy anesthetized patients to ensure a constant CI while assessing reliability. However, choosing this ‘ideal’ population may limit the generalizability of our findings. The reliability of the USCOM may be different in more unwell children with active cardiovascular pathology. Further examination of the USCOM in the less controlled, acute care setting will be needed. Also, we ensured the technical competency of our two observers by training them on 20 patients each prior to the study. This may also limit the generalizability. However, because of logistical delays, the study was conducted over 2 years, which reduced the impact of the initial training and is more reflective of clinical practice. Reliability was not affected over time. Our study proposes that with proper training precise measurements of the aortic outflow tract flow can be made. Appropriate initial and maintenance training with this tool is likely an essential part of ensuring user reliability.

Conclusions
This study confirms that the USCOM is relatively easy to use and reliable in healthy children when operated by trained users. It has the potential to be a valuable clinical and research tool in the assessment and management of cardiac output in children. Further study is warranted.

Acknowledgments
The authors acknowledge the subjects and their families for participation. Financial support was received through a new investigator grant from the Children’s Hospital of Eastern Ontario Research Institute. We thank the Clinical Research Unit and the Research Ethics Board at the Children’s Hospital of Eastern Ontario, and especially the operating room staff and the study coordinators involved. We also thank USCOM Ltd, Sydney for their equipment and technical support.

Disclosures
None of the authors have any financial or ethical conflicts of interest regarding the contents of this submission to disclose.
References


