LETTER TO THE EDITORS
Letters to the Editors are welcome and encouraged. Letters must raise points of current interest or address topics that have been previously addressed in the American Journal of Critical Care. Keep your letter concise. Letters are subject to editing. Include your name, credentials, title (optional), city and state, and telephone number (for verification, not for publication). Address to Kathleen Dracup, RN, DNSc, School of Nursing, University of California at Los Angeles, Factor Building, Box 956918, Los Angeles, CA 90095-6918, or fax, (310) 794-7482, or e-mail, AJCC@sonnet.ucla.edu.

To the Editors:

The recent article by Iregui and colleagues titled “Physicians’ Estimates of Cardiac Index and Intravascular Volume Based on Clinical Assessment Versus Transesophageal Doppler Measurements Obtained by Critical Care Nurses” (July 2003:336-342) deserves comment and very cautious interpretation because of multiple concerns with the study design and statistical analysis.

Accuracy of hemodynamic measurements obtained from transesophageal Doppler (TED) imaging is dependent on training and experience, and error can be easily introduced by inexperienced operators. Although nurses in this study received 2 hours of instruction about TED imaging and were supervised by investigators until they demonstrated proficiency, the authors failed to provide objective data to substantiate that proficiency in measuring hemodynamic variables with TED imaging was adequate before the study began. The authors cited research that validated specific objective training criteria to assess operator proficiency in obtaining cardiac output using TED imaging to ensure adequate reliability, but did not incorporate such criteria into their study design.

Nurses making hemodynamic measurements using TED imaging were blinded to physicians’ clinical estimates; they, however, were not blinded to pulmonary artery (PA) catheter data, current hemodynamic status, or therapy, all of which are well-known sources for introducing bias into measurements made with conventional transesophageal echocardiography. Seven nurses determined hemodynamic measurements using TED imaging during the study (mean of 15 procedures per nurse), and Doppler flow velocity waveforms were adequate in 99.1% of subjects. This information, however, does not provide data of intranurse and internurse variability in measuring cardiac output with TED imaging, which has not been established in previous studies that reported estimates of intraobserver and interobserver variability in measuring cardiac output with TED imaging. but elected not to incorporate them into their study design. Although training involving approximately 12 patients has been found to ensure reliability of cardiac output measurement with TED imaging, this estimate has not been established in nurses who are novices at TED imaging. Not accounting for intranurse and internurse variability casts serious doubt on the validity and reliability of study results for goals 1 and 3 as well as the authors’ conclusions.

Hemodynamic measurements made by novice nurses using TED imaging and measurements obtained with PA catheters (goal 3) should have been compared first to establish whether sufficient agreement existed between methods. These data were essential to establish whether the comparison of hemodynamic measurements made by nurses using TED imaging versus physicians’ clinical estimates (goal 1) was valid. The authors did not state that goal 3 was achieved before goal 1 in their study.

Precision in studies comparing measurement methods is affected by the variability in timing between paired measurements. Cardiac output varies significantly and spontaneously during brief periods (7-30 minutes) in critically ill patients receiving mechanical ventilation. The authors stated that measurements of cardiac index and estimates of IVV by nurses using TED imaging were made “within one hour” of physicians’ clinical estimates. No data were provided as to the actual time that elapsed between the physician’s estimate and TED measurement or whether treatment was added, altered, or deleted during this time. Cardiac index and/or IVV could have been influenced by midazolam (administered during placement of the Doppler probe and after the physician’s estimate), changes in mechanical ventilatory settings (eg, tidal volume, positive end-expiratory pressure), and titrations of vasoactive infusions. Failure to limit and control the time between physicians’ clinical estimates and TED measurements casts serious doubt on the validity of the reported data for goal 1.

The authors concluded that TED imaging can provide potentially useful estimates of cardiac output and IVV in critically ill patients. This conclusion was based solely on significant correlations found between the 2 methods among approximately 25 paired measurements. Simple linear regression is inadequate when comparing measurement methods because correlation...
coefficients (r values) are insensitive to systematic error, whereby high correlation coefficients can be attained even in the presence of significant measurement error.\textsuperscript{7,8,10} Correlation coefficients serve only to determine the level of association between 2 measurement methods, not the level of agreement. Hence, the data presented for goals 1 and 3 are inadequate to support the authors’ conclusions.

The level of agreement between 2 measurement methods is established by using the statistics of bias, precision, limits of agreement, confidence limits, and Bland-Altman plots, which allow for visualization of variability patterns across the measurement range.\textsuperscript{9-11} The authors elected not to incorporate these statistics into this study’s data analysis plan although they cited multiple research studies\textsuperscript{4,5,12} that employed these statistics and plots in determining agreement of cardiac output using the methods of TED imaging and PA catheters.

Because of these multiple scientific concerns, 2 conclusions of this study warrant very cautious interpretation by readers: (1) TED imaging can provide potentially useful estimates of cardiac index and IVV in critically ill patients who require mechanical ventilation (goal 3) and (2) physicians’ clinical estimates of cardiac index and IVV are often inaccurate (goal 1). This study did not provide valid or sufficient evidence to conclude that hemodynamic measurements made by nurses using TED imaging are comparable with hemodynamic measurements from PA catheters in critically ill patients. Future method comparison research should employ rigorous study designs and Bland-Altman statistics to establish whether hemodynamic measurements made by nurses using TED imaging can replace invasive measurements from PA catheters in guiding decision making in the care of critically ill patients.

REFERENCES


Nancy L. Szaflarski, RN, PhD
Burlingame, Calif

EDITORS’ NOTE: Iregui and colleagues declined to respond to this letter.

Continued on page 169

CORRECTIONS

Regarding the article titled “The Significance of Hypothermia in Preserving Ischemic Myocardium” (\textit{Am J Crit Care.} 2004;13:79-84), the title of the article appeared correctly in the article itself but contained a typographical error (correct spelling is “Hypothermia”) in the Table of Contents and on the cover. We regret the error.

On page 55 of the article titled “B-Type Natriuretic Peptide: A Diagnostic, Prognostic, and Therapeutic Tool in Heart Failure” (\textit{Am J Crit Care.} 2004;13:46-55), an error appeared in the CE test, question 9. It should have read as follows:

9. Which of the following is the recommended dose of nesiritide (Natrecor), as a therapeutic agent for use in heart failure, after an initial intravenous bolus?
   a. 0.01 µg/kg per minute
   b. 0.1 µg/kg per minute
   c. 1.0 µg/kg per minute
   d. 10 µg/kg per minute

The entire CE test is reprinted on page 171 of this issue of the Journal.
Letters to the Editors
Continued from page 101

To the Editors:
The article titled “Evaluation of Chemical Dot Thermometers in Orally Intubated Patients” (September 2003:403-408) by Potter and colleagues deserves comment and very cautious interpretation due to serious concerns with the statistical analysis.

The objective of this study was to “determine the accuracy of single-use chemical dot thermometers in orally intubated adults.” Accuracy is a term reserved for studies comparing a new measurement method with the gold standard method that provides the true value of the quantity being measured.1,2 When studies compare a new method with a clinical standard method (eg, electronic thermometers), accuracy cannot be determined since the true value of the measured variable remains unknown. Instead, the level of “agreement” is determined since each method contains its own level of measurement error.1,2 Although some may view this as semantics, standardization of terms and statistical analyses used in method comparison research has been called for recently to enhance quality comparison among method comparison studies and strengthen the methodology of future meta-analyses.2-4 Restated, this study’s true objective was to determine the level of “agreement” between single-use chemical dot and electronic thermometers in measuring oral temperature of intubated critically ill adults.

The authors reported that “almost 25% of all chemical dot thermometer readings were overestimates (11.8%) or underestimates (10.8%) by 0.4°C or more relative to the electronic readings.” This analysis was incorrect based on the authors’ a priori determination that “0.3°C temperature difference was determined to be clinically significant” (ie, 0.3°C and greater). This statement should read “approximately 40% of all chemical dot thermometer readings were overestimates (24.7%) or underestimates (16.5%) by 0.3°C or more relative to the electronic readings.” These corrected data provide initial descriptive evidence that agreement between the 2 methods was poor. The authors also do not report the distribution of differences between the methods in relation to the measurement range (°C). This critical information is needed to evaluate at what specific temperatures the greatest differences between methods occurred. Bland-Altman plots have become a standard feature of method comparison studies because they allow readers to visualize patterns that may exist between the variability in the differences of the 2 methods and the best estimate of the true value across the measurement range.2,5

The authors substantiate their conclusion that chemical dot thermometers are useful and reliable screening tools for orally intubated patients on the basis of the high correlation between the 2 methods ($r = 0.937$, $P = .01$), the lack of significant measurement difference between the 2 methods as determined by paired $t$ test, and a very low bias value (0.003°C). Correlation coefficients are insensitive to systematic error and do not provide actual estimates of agreement between methods.1,6 Significant high values of correlation coefficients can be derived if one method gives exactly twice the value of the other, and thus high correlation coefficients do not provide any confidence that methods agree. Correlation coefficients determine only the level of “association” between 2 measurement methods, not the level of “agreement.”1,6 Lack of significant measurement difference between methods determined by $t$ tests can result if the range of variability in the data for each method is broad, resulting in widely dispersed distributions that can lead researchers into falsely concluding that both methods are interchangeable.8 Hypothesis testing ($P$ values) does not determine the size of the measurement difference between methods and thus is inadequate to discern the level of agreement between measurement methods.

A conclusion about interchangeability of measurement methods cannot be based on bias alone because precision (random error) can also greatly affect agreement.1,2 The calculation of upper and lower limits of agreement (LOA) incorporates the standard deviation (SD) of the difference between methods (precision) to determine where 95% of the differences between methods lie relative to the mean difference, assuming a normal distribution. Values of the upper and lower LOA are then interpreted in context with a measurement difference determined beforehand, which defines clinical significance for the study.1,2,8 If LOA are smaller than this measurement difference, agreement between the 2 methods is sufficient and the 2 methods can be used interchangeably. Although the authors defined a measurement difference beforehand that represented clinical significance (0.3°C), they failed to report and evaluate LOA in their study. Hence, the study data do not substantiate the authors’ conclusion that chemical dot thermometers are useful and reliable screening tools in orally intubated patients.

The distribution of the mean differences (given in Table 2) was assumed to be normal by this author since bias had been calculated in this study (a statistic that assumes normality). This author calculated the LOA.
[bias (1.96 (precision))] for the differences between methods as follows: upper LOA = +0.62 [0.001 + 1.96 (0.318)] and lower LOA = -0.62 [0.001 - 1.96 (0.318)]. Since these upper and lower LOA (+0.62 to -0.62°C) exceeded ±0.3°C (clinical significant difference), this author concludes that temperature measurements from chemical dot and electronic thermometers did not sufficiently agree in this study and thus chemical dot thermometers are not interchangeable with electronic thermometers. Readers should note that this conclusion is the exact opposite of the authors’ conclusion.

In this era of patient safety, clinicians should incorporate fail-safe, accurate measurement methods into best practices. The measurement limits of chemical dot thermometers are 35.5°C to 40.4°C, which predisposes them to potential error in diagnosing states of hypothermia and hyperthermia in critically ill patients. Innovations in temperature measurement may provide better alternatives for practice such as the analog nonmercury glass thermometer designed for oral temperature measurement and available in the United States at approximately $4 to $6. These German-made thermometers contain a nontoxic alloy (Galinstan) composed of gallium, indium, and tin that measures temperature within one tenth of a degree in the range of 35.0°C to 42.0°C.9 Initial US studies have found bias and precision estimates of 0.11°C ± 0.19°C when Galinstan-glass thermometers were compared with mercury thermometers in nonintubated, hospitalized adults.10 High-quality method comparison research is warranted to discern if Galinstan-glass thermometers can be interchanged with electronic thermometers to provide a safer alternative for oral temperature measurement in intubated critically ill patients.

REFERENCES

Nancy L. Szaflarski, RN, PhD
Burlingame, Calif

The authors reply:

We appreciate the comments regarding our recent article, titled “Evaluation of Chemical Dot Thermometers in Orally Intubated Patients” (September 2003: 403-408). We regret the inconsistency caused by our reference to clinically significant differences of 0.3°C in discussing a priori power and ±0.4°C in discussing the results of Erickson et al2 as well as our own. Our intent was to focus on the latter definition of clinical significance.

The inclusion of limits of agreement (LOA) data would have enhanced our discussion of bias. Examination of the LOA with a Bland-Altman plot with LOA lines at ±0.62°C reveals only 3 cases (3.5%) falling outside the 95% LOA bounds and no apparent trend in extreme values in relation to estimated true temperature (average of chemical dot and electronic readings). Several other comparisons of bias in relation to the average temperature also revealed no trend. These included the computation of the Spearman correlation of the absolute value of the bias with the average temperature1 (r = -0.06) and a comparison of means of standard deviations for both bias and absolute bias for each of the 3 ranges of average temperature (35°C-36.99°C, 37°C-38.99°C, and ≥39°C). Bias data showed only small deviations from normality, and, as Bland and Altman indicated, “...a non-normal distribution of differences may not be as serious here [in LOA analysis] as in other statistical contexts.”

What is most salient is that the 95% LOA bounds (±0.062°C) exceed our clinically significant differences (±0.04°C). This negates a claim of equivalence; however, our results and discussion do not make such a claim. Our discussion regarding possible temperature underestimation with the chemical dot thermometers, our recommendation of their use as a screening tool with subsequent confirmation with an electronic thermometer when temperature measurements have important consequences for treatment decision making, and our caveat regarding the potential for user variability are appropriate cautions.

REFERENCES

Patricia Potter, RN, PhD, CMAC
Marilyn Schallom, RN, MSN, CCRN, CCNS
Susan Davis, RN, MSN, BC, CS-MSCNS, CCRN
Carrie Sona, RN, MSN, CCRN, CS-MSCNS
Maryellen McSweeney, PhD
LETTER TO THE EDITORS
Nancy L. Szaflarski

Am J Crit Care 2004;13 100-101
Copyright © 2004 by the American Association of Critical-Care Nurses
Published online http://ajcc.aacnjournals.org/

Personal use only. For copyright permission information:
http://ajcc.aacnjournals.org/cgi/external_ref?link_type=PERMISSIONDIRECT

Subscription Information
http://ajcc.aacnjournals.org/subscriptions/

Information for authors
http://ajcc.aacnjournals.org/misc/ifora.xhtml

Submit a manuscript
http://www.editorialmanager.com/ajcc

Email alerts
http://ajcc.aacnjournals.org/subscriptions/etoc.xhtml