Bias and Precision Statistics: Should We Still Adhere to the 30% Benchmark for Cardiac Output Monitor Validation Studies?

To the Editor:

It was with great interest that I read Peyton and Chong’s meta-analysis of papers comparing minimally invasive cardiac output monitors to thermodilution, particularly as much of the discussion centered around a much-cited paper by me and my late brother Julian Critchley. Our 1999 paper made a number of important contributions to assessing cardiac output monitoring: (1) it provided guidelines on how the results from validation studies should be presented; (2) it defined a standard statistical variable by which performance could be compared, the percentage error or limits of agreement; (3) it established that the error of the reference method, usually thermodilution, was also an important factor in validation studies; and (4) it set limits of acceptance of ±20% based on a precision for the reference method of ±20%. Peyton and Chong should be congratulated on providing a thorough and comprehensive meta-analysis of cardiac output validation studies published during the last decade. However, I would like to take issue with two of their conclusions.

Peyton and Chong make reference to the accuracy of thermodilution cardiac output. I would agree that the true precision of the thermodilution method in many of these validation studies is unknown. Cecconi et al. have suggested that the precision of the reference method should be known before using Bland and Altman analysis in validation studies. However, this is not easily done, and simply calculating the coefficient of variability from a series of readings is not the answer. The precision of the thermodilution can vary quite considerably depending of the type of catheter and measurement system, as we have shown in a recent in vitro investigation. However, I would have to take issue with Peyton and Chong in their assertion that precision is significantly worse during in vivo or clinical testing. I agree that cardiac output varies with respiration and other physiologic effects, but comparative measurements are made simultaneously, and the mean of several readings is used to average out these background physiologic effects.

Peyton and Chong also suggest that our ±30% limits of agreement are set too low and a more realistic benchmark would be ±45%, which they base on reference and test precision errors of ±30%, reset from our ±20%. Good news for the manufacturers. However, our choice of ±20% error for the reference readings was not just based on data from papers by Stetz et al. and Mackenzie et al., as Peyton and Chong suggest. In Bland and Altman’s 1986 paper, the decision as to what were acceptable limits of agreement was left to the judgment of the clinician, and for cardiac output measurement, most authors at the time were suggesting a precision of less than 1 l/min, which for a mean cardiac output of 5 l/min was 20%. Furthermore, increasing the limits ±45% would give a false impression that many monitors are measuring cardiac output reliably, which simply is not true. Therefore, I would still recommend using ±20% for our criteria.

Lester A.H. Critchley, M.D., F.F.A.R.C.S.I., F.H.K.A.M., Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong. hcritchley@cuhk.edu.hk

References


( Accepted for publication February 8, 2011 )

In Reply:

We appreciate the interest that Dr. Critchley has taken in our study and thank him for his letter. We agree that the real accuracy and precision of thermodilution are crucial to the discussion. Although thermodilution continues to be chosen as the clinical reference standard for validation studies, we would encourage further research to determine its precision, using high-precision tools, such as indwelling flow probes that can be adapted for use during major surgery and critical care. The evidence from recent studies using such invasive

Anesthesiology 2011; 114:1243–55

Copyright © by the American Society of Anesthesiologists. Unauthorized reproduction of this article is prohibited.
gold standards is that thermodilution has no better precision than the other methods surveyed during conditions of unstable hemodynamics.\(^1\)\(^2\) Furthermore, we welcome continued work by developers to improve the performance of new devices, and their subsequent independent testing in a variety of clinical scenarios.

Dr. Critchley’s comments confirm that the ± 20% criterion for agreement with the true cardiac output is essentially an arbitrary one. Our data suggest the likely limits of agreement of each generic method with the true cardiac output are closer to ± 30%. We leave judgment of the acceptability of this for clinical decision-making to the interested clinician. However, the studies quoted above suggest it is likely that this is the real precision of thermodilution that we have been routinely working with for many years, while managing patients during cardiac surgery.

Philip J. Peyton, M.D., M.B.B.S., F.A.N.Z.C.A.,* Simon W. Chong, M.B.B.S.* Austin Hospital and University of Melbourne, Melbourne, Australia. phil.peyton@austin.og.au

References


(Accepted for publication February 8, 2011.)

Reduction of Postoperative Mortality: Pattern of Use of β-Blockade, Bias, or Both?

To the Editor:

Wallace et al. reported the effects of the implementation of a hospital protocol for the addition or continuation of perioperative β-blockade in almost 40,000 patients at risk for myocardial ischemia and operated on between 1996 and 2008.\(^1\) The addition of perioperative β-blockade in eligible patients was associated with a significant reduction in 30-day and 1-yr mortality. Continuation of existing β-blockade in these patients also was beneficial, whereas withdrawal was reported to be associated with increased mortality (almost 400% increased 30-day mortality and almost 200% increased 1-yr mortality). Wallace and coworkers should be commended for this important and large study, which seems to confirm existing evidence about the continuation and withdrawal of perioperative β-blockade using “real world” data.

However, as also acknowledged by the authors, confounding by indication and selection bias are likely to have influenced the results of this retrospective analysis considerably. Therefore, we have some important questions regarding this study.

First, the authors tried to adjust for these potential sources of bias by collecting confounders and performing a propensity analysis. However, the logistic regression model described in table 5 of the article seems to include only previous coronary artery disease and peripheral vascular disease and not age, sex, and other potential confounders mentioned in table 4 of the article. We would like to see a table with the β-blockade effect measures adjusted for all potential confounders, because it is unclear whether these have been taken into account in table 5. Moreover, the methods of the propensity analysis are poorly described, which makes it difficult to interpret the results of these analyses.

Second, figure 3 of the article shows a markedly decreased mortality rate over time, which hardly can be attributed solely to the β-blockade protocol. It may also reflect a change in other practice patterns over time. This problem with retrospective studies with long duration is also recognized in the accompanying editorial.\(^2\) In this case, clonidine was added to the protocol in 2004.\(^1\) Likely, however, other drugs, such as statins or aspirin, were continued or prescribed more often as well in more recent years. Furthermore, there may have been improvements in surgical care, such as an increase in minimally invasive surgery. Apparently, these variables were not available to adjust for as confounders. However, adding “time” to the multivariable analysis as a proxy for a change in these variables may partially adjust the β-blockade effect measures for this potential confounding and could at least have been conducted as a sensitivity analysis.

Finally, previous comparable studies showed that including nadir and postoperative hemoglobin both in regression analysis and propensity analysis significantly influenced the β-blockade effect measures.\(^3\)\(^4\) If available, including these hemoglobin values in the analyses may therefore reduce the remarkably strong reported association between β-blockade withdrawal and outcome (odds ratio, 3.9; 95% CI, 2.6–6.0).

In conclusion, we would like to see the results of a regression model that includes both the variable of interest (pattern of β-blockade use) as well as all potential confounders, including time and, if available, both nadir and postoperative hemoglobin values, in a proper and crystal-clear analysis.

Wilton A. van Klei, M.D., Ph.D.* Linda M. Peelen, Ph.D., Wolfgang F. Buhre, M.D., Ph.D.* University Medical Center Utrecht, Utrecht, The Netherlands. w.a.vanklei@umcutrecht.nl

References