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,1 particularly as much of the discussion centered around a much-cited paper by me and my late brother Julian Critchley.2 Our 1999 paper made a number of important contributions to assessing cardiac output monitoring2: (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 ± 30% 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.1 However, I would like to take issue with two of their conclusions.

Peyton and Chong make reference to the accuracy of thermodilution cardiac output.3 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.3 However, this is not easily done, and simply calculating the coefficient of variability from a series of readings is not the answer.4 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.5 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.,6,7 as Peyton and Chong suggest. In Bland and Altman’s 1986 paper,8 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%.9,10 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

(Received for publication February 8, 2011.)

In Reply:
We appreciate the interest that Dr. Critchley has taken in our study1 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