Journal of Critical Care

VOL V, NO 2 JUNE 1990

EDITORIAL

Oxygen Kinetics: Pitfalls in Clinical Research

SOME INVESTIGATORS have reported that oxygen consumption in adult respiratory distress syndrome (ARDS) patients is uniquely dependent on delivery, regardless of the level of delivery. Other investigators report the usual biphasic relationship. In this issue of the *Journal of Critical Care,* there is a report by Fenwick et al[1] which purports to show that ARDS patients with elevated plasma lactate are on the supply-dependent slope of the O_2 delivery consumption curve, whereas those with normal lactate are on the independent portion of the curve. We would like to believe this study because it would make some sense out of the conflicting reports in the literature. Patients who are supply-dependent should show a correlation between delivery and consumption and should have lactic acidosis depending on the amount of time and the extent to which anerobic metabolism has occurred. Patients who are supply-independent should have no correlation between delivery and consumption, and no lactic acidosis. Therefore, the results and conclusions in this study fit our preconceptions of the way physiologic events should proceed. However, this study includes one important data point which must be in error, prompting us to examine the methodology more carefully.

The study of oxygen kinetics in critically ill patients is difficult because of unstable or fluctuating physiologic conditions; the use of drugs, volume expanders, and positive pressure ventilation (which affect the variables under study); the normal range of error in the measurement of cardiac output, hemoglobin, saturation, and oxygen consumption; the mathematical coupling of the allegedly independent variables in some studies; and the tendency to make statistical analysis and draw conclusions from as little as two data points per patient. The paper by Fenwick et al[1] exemplifies all of those potential methodologic problems and, hence, is worthy of careful study and discussion.

There are four major categories of potential error in conducting and analyzing this type of study:

1. accuracy of the primary measurements,
2. mathematical coupling of consumption and delivery calculations,
3. statistical analysis of complex interrelationships based on as few as two data points per patient, and
4. the definitions of "baseline," "steady state," and "pathologic physiology" in critically ill patients.

The accuracy and reproducibility of thermodilution cardiac output is the range of ±10%. This can be improved to ±5% with the use of a large volume of iced injectate with standardized injection guns. At best, the accuracy of hemoglobin measurement is ±2% and the accuracy of saturation measurement is ±2%. Thus, the error for calculating oxygen content is ±4% for both arterial and venous blood, and the potential error for calculating arteriovenous oxygen content difference is ±8%. The potential error for calculating systemic oxygen delivery as arterial content times cardiac output is in the range of ±10%. As pointed out by Stratton et al,[2] this potential for error is insignificant compared with the 100% to 200% variation in systemic oxygen delivery, which is reported in most experiments. However, the range of error for calculated oxygen consumption (arteriovenous O_2 difference times cardiac output) is in the range of ±15% for each data point.

This must be taken into account when investigators claim that a 16% change in oxygen consumption, in response to an intervention, is significant (as claimed in this paper). Of course, we all assume that a variation in measurement will occur in a random fashion, making it acceptable to make calculations with specific numbers rather than a range. However, when there is only one data point before and one data point after an intervention, and if all of the variations in each of the measurements were on the low side before the intervention and on the high side after the intervention, calculated oxygen consumption would appear to increase by 30% for that specific patient in response to the intervention when, in fact, there may be no difference at all. This problem is common to all physiologic investigations, but, in this case, there is a partial solution. Oxygen consumption can be measured directly across the airway with an accuracy and reproducibility of ±5%. Since that technology is readily available, it makes sense to minimize the potential errors by measuring oxygen consumption directly.

Most of the studies of oxygen kinetics do not involve direct measurement of oxygen consumption, but they do calculate both oxygen consumption and oxygen delivery from the same two or three measurements (arterial content, venous content, and cardiac output). This practice not only introduces the potential for significant calculation errors of VO_2 (as previously discussed), but mathematically couples the two variables we would like to evaluate. If Fick[3] was right, this practice should be totally acceptable. Since arterial blood is almost fully saturated in most circumstances, the discussion narrows down to the interrelationship between cardiac output and venous oxygen content. If metabolic rate (VO_2) remains constant, an increase in cardiac output should be accompanied by an exactly proportionate increase in venous content, and vice versa. At constant cardiac output, an increase in VO_2 should be accompanied by an exactly proportionate decrease in venous content, and so on for every possible combination of VO_2, cardiac output, and venous content. These relationships should be true whether consumption and delivery are measured independently or whether they are calculated from the same primary measurements. The problem comes in actual application. Suppose we are trying to get a data point on a critically ill septic patient with ARDS on a mechanical ventilator and several drugs which alter cardiac performance and metabolic rate. We do five sequential thermodilution injections to maximize accuracy, with results ranging from 4.2 to 5.0 L/min/m^2. It takes 10 minutes to do the five thermal curves. During that time, the continuous readout of venous saturation varies from 66% to 72%, fluctuating with each respiratory cycle. When do we draw the venous blood for analysis? Surely not with each thermodilution injection. How do we match the hemoglobin and the venous content to the thermal curves? Should we select the curve done closest to the time of venous sampling, or shall we select the average of the five curves? Suppose the venous saturation is measured in an oximeter at 63%. This was never observed on our pulmonary artery continuous monitor. Which oximeter is more accurate? We should really measure oxygen content directly, but the Lex-O-Con is broken and the technician who knew how to run the Van Slyke and Scholander apparatuses has long since retired. We therefore settle on the average thermodilution output and the in vitro measured saturation and we have a data point. Now we give 2 *U* of packed red blood cells and return in a few hours to go through the same process, hoping that all the measurement variations will cancel each other out when we have developed a large enough population. However, for that particular patient, there are only two data points, one before and one after the intervention. Fick[3] was right on paper, but he never tried to do these experiments. We would all feel more confident if the measurements of consumption and delivery were not calculated from the same primary data.

When two physiologic variables have a complex relationship it would seem wise to develop as many comparative data points as possible. Measurement of blood glucose before and 1 hour after glucose ingestion is a helpful screening test but not a glucose tolerance curve. Would we accept a study describing the effect of preload on cardiac output in 20 critically ill patients based on two data points per patient, each representing some snippet of full Starling curve? Unlikely. We should then be suspicious of any physiologic study of two related variables which does not include many data points for each patient cover-

ing the major portion of the curve. This is easy to do in the laboratory, but much more difficult to do in the clinical setting. Difficult, but not impossible.

Just contemplating the problems of generalizing on two data points should make us worry. In the study by Fenwick et al, there is one patient in group B who is readily identifiable as an outlier simply by reviewing Fig 1, bottom. This patient's oxygen consumption index prior to transfusion is approximately 30 mL/min/m^2 and rises to 200 mL/min/m^2 after transfusion. The normal range of oxygen consumption is 80 to 100 mL/min/m^2. At a normal level of oxygen delivery (which this patient has), a metabolic rate one third of normal would occur only in a patient who is very hypothermic or has sustained some mitochondrial poisoning. This data point must represent an error in measurement or, at least, some radical change in the steady state before and after transfusion.

The elimination of the erroneous data point and its matched pair would not dramatically alter the significance observed when using the Wilcoxon sign rank sum test. Some discussion is warranted concerning the methodology of statistical analysis used by Fenwick et al. A common method for parametric analysis of paired data is the paired *t* test. This test assumes an underlying normal distribution. It is considered a robust test because it can accommodate some deviation from normality. However, due to the small sample size in this experiment (n = 11 in group A and n = 13 in group B), the investigators correctly selected a distribution-free analysis method. The null hypothesis of this method states that there is no difference in rank sums between pretransfusion and posttransfusion values. As demonstrated by the investigators, the null hypothesis can be rejected at a critical value less than or equal to 0.01. In other words, it is unlikely that the observed difference occurred by chance. Unfortunately, the investigators concluded that their data demonstrates a pathologic *dependence* of oxygen consumption on oxygen delivery. This conclusion is not consistent with the analysis used. A more appropriate analysis method for test of dependence would be the Pearson product moment correlation coefficient (r^2). Correlation of the change (actual or relative) in both indices would seem appropriate for this experiment. If the investigators had examined the correlation coefficients between VO_2 and DO_2 for both groups, it is unlikely that they would have reached the same conclusion. We found no significant difference in correlation coefficients between consumption and delivery in either group (r^2 approximately .26 and .16 for groups A and B, respectively), indicating that a transfusion-induced increase in delivery is not associated with a concomitant rise in consumption, whether the lactate is elevated or not.

Studies like this are more complete if they include several data points covering a wider range of O_2 delivery.

Following the analogy of preload versus cardiac output, we would expect that each patient in such a study would be functioning on a different "Starling curve," and a single patient may move from one Starling curve to another during the course of the study in response to the intervention, or in response to some other variable. We would not expect to determine a specific critical level of left atrial pressure below which cardiac output is inadequate in such a group of patients. Rather, we would expect to find a wide range of response between preload and cardiac output, with the full spectrum of response demonstrated only when the full curve is drawn out for each patient and each situation. In the study of oxygen kinetics in the critically ill, we should expect the resting VO_2 to be between 100 and 240 mL/min/m^2; the more septic the patient, the higher the metabolic rate. The normal response to this increased metabolic rate is a compensatory increase in systemic oxygen delivery effected by an endogenously mediated increase in cardiac output. Just as in exercise, anemia, or hypoxia, cardiac output will increase until the ratio of delivery to consumption is reestablished at the normal level of approximately 5:1. If the cardiac response to hypermetabolism is limited by cardiac disease, hypovolemia, or increased intrathoracic pressure, the patient will stabilize at some lower ratio of delivery to consumption (the coefficient of oxygen extraction will be higher). When the ratio falls to some level below 2:1, we should expect that the patient will be supply dependent with all of the attendant associations. The critical level of oxygen delivery at the 2:1 ratio will be approximately 330 mL/min/m^2 if the metabolic rate is 160 mL/min/m^2, but the "critical" level

may be much higher if the patient is hypermetabolic. In any group of critically ill patients, we should expect to find some who are supply dependent, some who are not, and some who are near the "knee" of the curve. It is incorrect to describe a supply-dependent state as "pathologic" simply because the oxygen delivery is less than 330 mL/min/m^2.

Given all of these potential problems, variability between and among patients, and the limitations to experiment posed by good patient care, how can we determine the relationships between consumption and delivery in critically ill patients and decide on the "optimal" level of oxygen delivery, which we should try to achieve? The following recommendations would be a good starting point:

1. Measure oxygen consumption and oxygen delivery independently.

2. Characterize the baseline steady state with at least three measurements taken 10 to 15 minutes apart. Get as many data points as possible during and after the intervention. Plan the study so that for each patient and each intervention there will be at least three data points above a 4:1 delivery consumption ratio to document the independent portion of the curve (if there is one), with data points extending to ratios below 2:1 to document the position of the knee of the curve (if there is one).

3. Calculate and analyze each measurement as the studies are done, so that data far outside the expected range can be either verified or disproven.

4. Resist the temptation to generalize until there are enough data to describe most of the curve for each patient at each clinical setting.

Fenwick et al may be correct in stating that elevated lactate detects the supply-dependent condition, but that hypothesis is not proven by the methods and data presented in their report.

Robert H. Bartlett, MD
Ronald E. Dechert, MS
University of Michigan Medical Center
Ann Arbor, MI

REFERENCES

1. Fenwick JC, Dodek PM, Ronco JJ, et al: Increased concentrations of plasma lactate predict pathologic dependence of oxygen consumption on oxygen delivery in patients with adult respiratory distress syndrome. J Crit Care (this issue)

2. Stratton HH, Feustel PJ, Newell JC: Regression of calculated variables in the presence of shared measurement error. J Appl Physiol 62:2083-2093, 1987

3. Fick A: Veber die Messung des Blutquantums in den Herzventrikeln. SB Phys Med July 9, 1870