EDITORIALS

Significance of Serial Echocardiographic Measurements

The reliability and significance of many echocardiographic results are controversial despite extensive literature describing echocardiographic findings and widespread clinical use of the technique. A major reason for this situation is the difficulty in standardizing both the performance of the examination and the interpretation of the records. The importance of reproducibility is frequently acknowledged, but very little work has been done in that area.

Pietro et al, in this issue of Chest (see page 29), report the reproducibility of serial echocardiographic measurements among clinically stable patients and compare it with intraobserver variability in echocardiographic measurements for the same patients. Intraobserver variability for both normal subject and cardiac patient groups was low and statistically nonsignificant. Variability in measurements from serial examinations was two to three times as great and was statistically significant, even though both studies were performed by the same technician, and care was taken to have the patient in the same position of recumbency and the transducer in the same location for both examinations. Variability was expressed as a percentage by dividing the absolute difference between measurements by the first measurement and multiplying the result by 100. Variability (reproducibility) described by this method is easy to understand and can be used by others in evaluating the relevance of serial differences in measurements for individual patients. Reports of variability in measurements should also include the range of differences and the confidence intervals.

Percentage of uncertainty (1.97 SD divided by the mean) has been used by others to describe variability in echocardiographic measurements, but it does not provide the information needed for assessing the significance of different measurements for an individual patient. Percentage of uncertainty as just defined for describing variability depends on normal distribution of data. This restriction can be removed by using the actual confidence intervals derived from the distributions.

The left ventricular internal dimension at end-diastole (LVD) showed the least intraobserver and reperformance variability in the study of Pietro et al. Results of this study indicate that a change in LVD of 5 percent or greater is not due to problems of reproducibility and provides a basis for determining statistically significant serial changes in clinical situations. Prakash reported no significant change in LVD or at end-systole among echograms repeated over a 90-minute period in uncomplicated acute myocardial infarction patients. A total of 825 dimensions were measured for 23 patients, and maximal variation among both diastolic and systolic dimensions was 3 mm for an individual patient, with an average of 1.4 mm. LV dimensions are also the most reproducible echocardiographic measurements in studies comparing interobserver variability. Intraobserver and interobserver variability for LV dimensions for patients examined at the Krannert Institute of Cardiology is $\leq 2$ mm, while reperformance variability among stable patients is $\leq 3$ mm. Expressed as a percentage, this variability is less than 5 percent.

Reproducibility of interventricular septal and posterior LV wall thickness measurements is poor compared with that for LV cavity dimensions. This is partially because percentage of variability is greater for small structures when the absolute difference is divided by a comparatively small denominator. Another reason for poor reproducibility of septal thickness measurements is the difficulty of identifying echoes from the right and left sides of the septum. Echoes from the right side of the septum are difficult to differentiate from echoes reflected from tricuspid chordae, papillary muscle, and trabeculations within the right ventricular cavity. Inclusion of these right-sided structures yields an inaccurate measurement of septal thickness and may give a false impression of septal hypertrophy. Echoes from the left side of the septum are often missed ("dropout"), which results in an erroneously small septal
thickness measurement. Reproducibility of posterior LV wall thickness measurements is also limited by difficulty in defining exactly which echoes are reflected from endocardium and epicardium and by the varying amounts of myocardial trabeculation. The poor correlation between LV wall thickness measurements on serial echocardiograms reported by Monoson et al. indicates a necessity for caution in ascribing quantitative changes in wall thickness and systolic thickening to stress or therapeutic interventions.

There are many reasons for differences in measurements between serial echocardiographic examinations. Some causes of variability are eliminated easily, and others are very difficult. Sahn et al. have made recommendations for standardization of measurement criteria that should reduce interobserver variability. Range resolution of current M-mode echocardiographs is approximately 1 mm, which means there is a potential error of about 10 percent in measuring LV wall thickness (normal range, 7 to 12 mm) owing to equipment limitations. Lateral resolution is not as good as range resolution, and "spurious echoes" are a common cause for error in structure identification. There is also random error because of variation in the velocity of sound traveling through a nonhomogeneous medium. Reverberations of sound may cause mistakes in interpretation, but those mistakes are usually avoidable. The importance of gain control settings recently has been emphasized by Martin et al. in a report showing how different gain settings can cause significant and clinically important differences in the two-dimensional echocardiographic determination of mitral valve orifice area. Variability in serial examinations can be reduced by using the same equipment, control settings, patient position, and transducer location for a given patient. Variations in measurement owing to respiration can be controlled by making all measurements from echoes recorded during expiration. The patient’s heart rate, blood pressure, and metabolic state influence cardiac measurements and must be considered in comparing serial echocardiographic tracings, although they cannot be adequately controlled.

The most difficult problems of reproducibility to eliminate are those of structure identification, particularly in patients who are difficult to examine echocardiographically. Technician expertise and thoroughness are critical in minimizing those errors because the physician interpreting the record sees only those echoes selected by the technician. Two-dimensional (2D) echocardiography is often helpful in understanding the origin of confusing echoes obtained by M-mode. However, the problems with resolution, echo dropout, and spurious echoes are as great or greater with 2D echo, and Pietro et al. found no improvement in reproducibility between M-mode measurements obtained with 2D echo guidance and those obtained by standard M-mode techniques.

The numerous factors influencing reproducibility make it imperative for each laboratory to test carefully reliability and variability for its own particular equipment and personnel. Differences in reproducibility between laboratories may be the cause of conflicting results in clinical studies that rely on echocardiographic measurements.

The idea that quantification by echocardiography is most useful when serially applied to the same patients is not clearly supported by results of the serial echocardiographic studies reported by Pietro et al and Monoson et al. The significant variability between serial studies makes it particularly difficult to evaluate the relatively small quantitative changes in echocardiographic measurements of cavity dimensions and wall thicknesses that occur with stress and drug or other therapeutic interventions. Research protocols for evaluating the effects of interventions, therefore, should include testing for reproducibility under the circumstances of the experiment.

Betty C. Corya, M.D.
Indianapolis

Krannert Institute of Cardiology; Department of Medicine, Indiana University School of Medicine; and Veterans Administration Medical Center.

Supported in part by the Herman C. Krannert Fund, National Heart, Lung, and Blood Institute grants HL 06308 and HL 07182, and by the American Heart Association, Indiana Affiliate, Inc.

Reprint requests: Dr. Corya, Indiana University School of Medicine, 1100 West Michigan, Indianapolis 46223

REFERENCES

6 Martin RF, Rakowski H, Kleiman JH, Beaver W, London E, Popp RL. Reliability and reproducibility of two dimen-

2 EDITORIALS

CHEST, 79: 1, JANUARY, 1981
Clinical Measurement of Lung Water Content

This report summarizes the recent workshop attended by 35 investigators held in Vancouver, BC, to assess the current status and future potential for the measurement of lung water content, particularly in terms of human applications for early detection of interstitial edema.

The chest x-ray is the principal method for the detection, quantification, and distribution of pulmonary edema. Rightly or wrongly, it is the standard with which all other methods are compared. The advantages of the x-ray are that it is relatively noninvasive, is inexpensive, can be done repeatedly (even under adverse conditions), and permits assessment of regional differences.

What it is not clear is whether the x-ray is sensitive enough to allow detection of small increases in lung water as in early interstitial edema or whether one can reliably tell the difference among vascular congestion, changes in lung volume, and interstitial edema. The sensitivity of the chest x-ray has been compared with the directly measured lung water in animals. When extravascular water was increased by more than 30 percent, the radiologist always saw changes.

Pulmonary function tests have never been considered specific enough to be of use in detecting pulmonary edema. Recent studies in animals indicate that increases in lung water in the range of 20 to 40 percent significantly reduce several pulmonary function measurements, including functional residual capacity, dynamic lung compliance, and arterial oxygen tension.

The inhaled soluble gas measurement of pulmonary tissue volume is being used by several groups. Most agree that a rebreathing procedure using acetylene is the method of choice. Tissue volume changes ought to be sensitive in early interstitial edema, before ventilation distribution is greatly affected. Comparison of tissue volume with the chest x-ray in patients with early congestive heart failure shows a good correlation on the whole, but in any one individual a good chest x-ray is easier to obtain and is probably more reliable.

Transthoracic electrical impedance does not appear to be useful. Clinical tests indicate that the normal range is large, and the sensitivity is low compared with the chest x-ray.

The double indicator dilution to measure extravascular lung water, once in vogue but largely abandoned because of its low accuracy and high technical requirements, appears to be making a comeback. One group, using heat as the diffusable indicator and on-line microprocessors for computation, report the measurement is now faster, easier, more reproducible and can be done repeatedly. Although the double indicator dilution measurements require placement of catheters in both the pulmonary artery and aorta, the thermal dilution approach is less demanding than the traditional one because all of the blood that is withdrawn is returned immediately to the subject.

Evidence from animal experiments indicates that there is an excellent correlation between the thermal dilution lung water and lung water measured directly by gravimetric methods. Clinical tests are in progress. The consensus is that the thermal dilution procedure, although having some unanswered theoretical questions, warrants continued intensive testing over a wide range of patient conditions, especially uneven blood flow distribution.

The exploding technology of nuclear medicine focused attention on several exotic methods to quantify lung water. These include lung density by gamma-ray attenuation or by Compton scatter, CT scanning, and nuclear magnetic resonance. All of these methods are theoretically sound, but none has yet achieved sufficient reliability or accuracy to warrant clinical applications. One disheartening feature of these approaches is that the cost increases exceedingly rapidly with the degree of technical sophistication.

This brings us to the crux of the matter; namely, what is the ultimate value of the measurement of lung water content? No matter how precise and reliable such methods become, they are static measurements limited, by their very nature, to detecting increased lung water content. Since the normal range of lung water content is wide, these methods work best when used sequentially to follow up a patient’s clinical course. Frequently, the beginning of acute lung injury does not cause a measurable increase in extravascular lung water. Therefore, measurement of lung water is probably not useful in early lung injury.