EDITORIAL COMMENT

How Reliable Are Serial Echocardiographic Measurements in Detecting Regression in Left Ventricular Hypertrophy and Changes in Function?*

Julius M. Gardin, MD
Orange, California

Left ventricular (LV) mass and geometry determined by echocardiography have convincingly been shown to predict cardiovascular morbidity and mortality independent of other cardiovascular disease (CVD) risk factors (1–7). For example, in the Framingham Heart Study, subjects ≥40 years old without clinically apparent CVD who were followed for four years demonstrated a risk factor–adjusted relative risk of death of 1.5 in men and 2 in women for each 50-g increment of echocardiographically determined LV mass adjusted for height (1). In addition, various reports have suggested that echocardiographic LV mass may be a suitable measure of subclinical disease, reflecting a stage in the process beyond, and the accumulated effects of traditional CVD risk factors (e.g., hypertension, hyperlipidemia, obesity) (8,9). Furthermore, echocardiography has been shown to be a more sensitive tool than electrocardiography for the detection of LV hypertrophy (10,11). Although the data are still sparse, regression of LV hypertrophy, as demonstrated by echocardiography, appears to be associated with reduced cardiovascular morbidity (12,13). Consequently, reliable estimation of LV mass has been an important goal for researchers and clinicians interested in topics such as CVD risk stratification, detection of subclinical disease and measurement of potential regression of LV hypertrophy related to treatments designed to reduce blood pressure and control obesity (14). In addition, measurements of LV function (e.g., global LV ejection fraction) have also been shown to be important indicators of cardiovascular prognosis (15).

Problems of measurement variability and reliability. The use of serial echocardiographic measurements of LV mass (e.g., in an attempt to detect the effect of antihypertensive therapy) has been hampered by problems of measurement variability and reliability (Table 1) (16–20). For example, Gottdiener et al. (16) reported on 96 patients with hypertension evaluated with two-dimensionally targeted M-mode echocardiography twice within 6 ± 8 days. Studies were performed by the same sonographer using the same echocardiography machine and were read by the same echocardiogram reader. Despite excellent measurement reliability (intraclass correlation coefficient [RHO] = 0.86), the 95% confidence interval width of a single replicate measurement of LV mass was 59 g, exceeding the usual decreases in LV mass observed during antihypertensive treatment. The test–retest reliability of LV mass measurements was highly influenced by study quality, which depended importantly on age and weight. The authors noted that in clinical practice, or in studies in which factors such as machine, sonographer and reader are not standardized, the 95% confidence intervals for LV mass measurements might be even greater than they observed.

Other sources of measurement variability and bias in the echocardiographic estimation of LV mass have included those related to interreader and intrareader variability, intersonographer and intrasonographer performance variability, biologic variability from beat–to–beat and day–to–day and missing measurements (unmeasurable LV mass). In the Coronary Artery Risk Development in Young Adults (CARDIA) study, measurement variabilities (technical errors) for two-dimensionally directed M-mode LV mass on the first echocardiographic examination were intrareader 8%, interreader 14%, intrasonographer 10% and intersonographer 10% (21). Similarly, in the Cardiovascular Health Study, a multicenter study of cardiovascular morbidity and mortality in over 5,000 free-living elderly subjects, the intrareader and interreader measurement variabilities (mean percent difference between measurements) for LV mass on the first echocardiographic examination were both 10% (20). In addition, in the Cardiovascular Health Study, two-dimensionally directed M-mode echocardiographic LV mass could not be adequately measured in approximately one-third of the subjects (20,22). Factors associated with missing LV mass measurements included obesity, diabetes, male gender, white race, history of myocardial infarction and presence of echocardiographic LV wall motion abnormalities. Clearly, the association of missing LV mass measurements with known CVD or CVD risk factors introduced selection bias as an issue, because patients in whom LV mass could be measured represented a lower risk subgroup of the entire cohort. One strategy adopted to overcome this bias was the development of an electrocardiographic model for LV mass estimation based on a comparison of the two-thirds of the Cardiovascular Health Study subjects in whom LV mass estimates could be derived from echocardiography as well as electrocardiography (23).

Another important source of variability in serial studies is

*Editorials published in the Journal of the American College of Cardiology reflect the views of the authors and do not necessarily represent the views of JACC or the American College of Cardiology.

From the Division of Cardiology, Department of Medicine, University of California at Irvine Medical Center, Orange, California.
Table 1. Echocardiographic Estimation of Left Ventricular Mass: Sources of Measurement Variability and Bias

<table>
<thead>
<tr>
<th>Source of Variability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Intrareader</td>
<td>Variability between measurements made by the same reader</td>
</tr>
<tr>
<td>2. Interreader</td>
<td>Variability between measurements made by different readers</td>
</tr>
<tr>
<td>3. Intrasonographer</td>
<td>Variability introduced by the sonographer</td>
</tr>
<tr>
<td>4. Intersonographer</td>
<td>Variability introduced by the scanner</td>
</tr>
<tr>
<td>5. Beat-to-beat variability</td>
<td>Variability due to changes in measurement between heartbeats</td>
</tr>
<tr>
<td>6. Biologic (day-to-day) variability</td>
<td>Variability due to changes in patient condition over time</td>
</tr>
<tr>
<td>7. Reader (temporal) drift</td>
<td>Variability due to changes in measurement over time</td>
</tr>
<tr>
<td>8. Regression to the mean</td>
<td>Variability due to regression to the mean phenomenon</td>
</tr>
<tr>
<td>9. Left ventricular mass unmeasurable (missing)</td>
<td>Cases where measurement could not be made</td>
</tr>
</tbody>
</table>

the potential for change in the exact method by which readers make measurements over time. For example, in the CARDIA study, the variability in measurements of LV mass made on the same studies measured five years apart by a single reader was in the range of 16%, as compared with the ~4% intrareader variability for LV mass measurements made on studies reread by the same reader within a few weeks of each other during the initial year. This temporal drift in measurements resulted in redesigning the protocol to reread the initial studies in temporal proximity to the follow-up studies recorded five years later (Gardin JM, unpublished data).

Important findings of the current study. In this issue of JACC, Palmieri et al. (24) report data on intrapatient reliability of LV mass measurements in 183 hypertensive subjects, all of whom were judged to have LV hypertrophy on their screening echocardiograms, from the Prospective Randomized Enalapril Study Evaluating Regression of Ventricular Enlargement (PRESERVE). A second echocardiogram was repeated at the time of randomization (45 ± 25 days later). Analyses of intrapatient reliability were reported for comparisons of measurements of LV mass, internal dimension and wall thickness, systolic function and diastolic filling. A strength of this report is that in addition to the two readings, the authors present comparison data, at least for LV mass, between the observed measurement difference and the expected difference based on a formula used to predict "regression to the mean," a statistical artifact that may confound attempts to detect serial measurement changes (25,26). Specifically, as the authors note, this regression phenomenon typically occurs in selected subjects (e.g., patients with LV hypertrophy) who have a variable (i.e., LV mass) that exceeds a specified partition value. Because of regression to the mean, repeat studies in the selected group of patients (with high values for LV mass) may reveal a lower mean value of LV mass on the second study simply because of random fluctuation of measurements. Parenthetically, regression to the mean may also be operative when one chooses patients with the lowest values in the group. The magnitude of regression to the mean (R[g], in grams) likely to affect the "change" in LV mass measurement over time can be estimated by the formula (25,26): R[g] = (H - N) × (1 - r), where H is the average value of LV mass in a patient group selected for high baseline values of LV mass; N is the center of distribution of LV mass within a reference group; and r is the interindividual variability/total variability for the LV measurement. Using a calculated reference value of 170 g for LV mass based on the report of Hammond et al. (27), as well as a ratio r of 0.70, Palmieri et al. (24) derive, as did Herpin and Demange (26), a simplified formula for predicting LV mass on the second (randomization) study from the following formula: R_g(%) = 3910/LVM_g - 23, where R_g(%) is the percent reduction expected for absolute LV mass at the second evaluation (based on regression to the mean) and LVM_g is LV mass measured on the initial study (in grams).

Of interest, in the Palmieri report, mean average LV mass decreased less between the first and second echocardiograms than would be expected from the above "regression to the mean" formula (2 ± 19 vs. 17 ± 12 g, p < 0.001). One factor that may have contributed to this relatively small difference was the small increase in blood pressure noted between the first and second echocardiograms, which may have resulted in some increase in LV mass, thereby blunting the expected regression to the mean. However, this confounding situation, if present, no doubt had a minor effect on the measurement difference. As emphasized by Palmieri et al. (24) and other investigators (28), the higher the reliability of the measurements, the less regression to the mean would be expected. In this regard, the authors employ careful methodoloy, including strategies described later, to decrease echocardiographic measurement variability. The excellent intrastudy reliability is reflected in high RHO for LV mass (0.93), LV internal diameter (0.87), ventricular septal thickness (0.85) and LV posterior wall thickness (0.83). In addition, substantial or moderate reliability was observed for measures of LV systolic function and diastolic filling, with RHO ranging from 0.71 to 0.57.

Specific measures used by the authors to decrease measurement variability included centralized reading of all studies with a single, highly experienced final arbiter of readings, use of a standardized echocardiographic protocol and a "hands-on" training program for sonographers (29). In addition, the authors substituted linear measurements of LV wall thickness and internal dimension from the two-dimensional parasternal long-axis view whenever the two-dimensionally directed M-mode beam was not ideally oriented for deriving thickness and dimension measurements. Because of the relatively older ages (all were at least 50 years old) and body mass indexes of the study participants, the majority of measurements in the study were made directly from the two-dimensional echocardiographic views. This strategy was no doubt useful in avoiding the selection bias related to missing LV mass measurements in elderly subjects at higher risk for CVD noted in the Cardiovascular Health Study (22). In addition, the design of the current study, in which both screening and randomization echocardiograms were recorded within a relatively short period (45 ± 25
resonance imaging has been reported in an in vivo canine perior accuracy for LV mass determination by magnetic dimensional echocardiographic image acquisition (33). Su-
localization of image slices (32) and real-time three-
of LV mass have been reported using three-dimensional over standard two-dimensional echocardiographic estimates.

gators agree on the superiority of two-dimensional echocardiography (31). However, not all investiga-
calculations performed from two-dimensionally directed echocardiographic slice areas and thicknesses to be superior to measurements calculated from two-dimensional echocardiographic studies. Some workers have reported LV mass in the face of excellent alternative—although currently more cumbersome and expensive, noninvasive imaging technologies such as magnetic resonance imaging and computed tomography.

strategies for minimizing measurement variability.

Strategies that have been proposed for minimizing measurement variability in echocardiographic studies involving measurement of LV mass, including serial studies designed to detect changes (regression) in LV mass, include those summarized in Table 2 (20,22,24,30).

In addition, other echocardiographic techniques (e.g., two-dimensional [31] or three-dimensional [32]) can be used to estimate LV mass. Some workers have reported LV measurements calculated from two-dimensional echocardiographic slice areas and thicknesses to be superior to calculations performed from two-dimensionally directed M-mode echocardiography (31). However, not all investigators agree on the superiority of two-dimensional echocardiographic estimates of LV mass. Further improvements over standard two-dimensional echocardiographic estimates of LV mass have been reported using three-dimensional localization of image slices (32) and real-time three-dimensional echocardiographic image acquisition (33). Superior accuracy for LV mass determination by magnetic resonance imaging has been reported in an in vivo canine study, with equivalent accuracy using end-diastolic ($r = 0.94, \text{SEE} = 9 \text{ g}$) or end-systolic ($r = 0.97, \text{SEE} = 7 \text{ g}$)
frames (34). In comparison, in an early necropsy comparison study, the standard error of two-dimensionally derived M-mode echocardiographic measurements for estimating LV mass was $=30 \text{ g}$ (35).

In summary, this study by Palmieri et al. (24) adds significantly to our understanding of the optimal methods for achieving the excellent measurement reliability necessary for application of echocardiographic measurements to de-

tecting regression in LV mass after appropriate treatment (for hypertension, obesity, etc.). This rigorous approach to improving reliability of measurements is critical if echocardiography is to be used to document serial changes in LV mass in the face of excellent alternative—including currently more cumbersome and expensive, noninvasive imaging technologies such as magnetic resonance imaging and computed tomography.

Reprint requests and correspondence: Dr. Julius M. Gardin, Division of Cardiology, University of California Irvine Medical Center, Rt. 81, 101 City Drive South, Building 53, Room 100, Orange, California 92868-4080.

REFERENCES
