INTER-OBSERVER VARIABILITY IN ECHOCARDIOGRAPHY INTERPRETATION

Adhikari PM¹, Chowta MN²*, Pai N³, Kamath P⁴, Kamath A⁵

¹ M.D, Professor of Medicine, Kasturba Medical College, Mangalore
² M.D, Associate Professor of Pharmacology, Kasturba Medical College, Mangalore
³ M.D, D.M. Assistant Professor of Medicine, Interventional Cardiologist. Kasturba Medical College, Mangalore
⁴ M.D, D.M. Associate professor of Medicine, Kasturba Medical College, Mangalore
⁵ MSc, Senior Grade Lecturer, Department of Medical Statistics, Kasturba Medical College, Manipal.

Correspondence author E-mail address: muktachowta@yahoo.co.in

Objective: The objective of this study is to evaluate the inter-observer variability in the echocardiography interpretation.

Settings and Design: Tertiary care hospital, Prospective study.

Subjects: A total of 37 patients were included in the study. Echocardiography was done in each patient four times at 2 weeks intervals by different cardiologists. Each cardiologist was blinded to the previous reports. Echocardiography reports of each patient were analyzed for any discrepancy in interpretation. The parameters used were left ventricular ejection fraction, presence of diastolic dysfunction, presence of pulmonary hypertension, regional wall motion abnormality, valve abnormality and overall interpretation.

Results: There was fair agreement with regard to diastolic dysfunction for reporter 1 and 2, reporter1 and 3, reporter1 and 4, reporter 2 and 3. Agreement was slight for reporter 2 and 4, reporter 3 and 4. Similarly for detection of valvular abnormality, there was only a slight agreement between reporter 1 and 2, reporter 2 and 3. Agreement between reporter1 and 3, reporter 1 and 4, reporter 2 and 4, reporter3 and 4 were fair for detection of valvular abnormality. Agreement for overall interpretation of echocardiography showed better correlation for reporter1 vs reporter 3 and reporter2 vs reporter3. Correlation for other pairs was slight. Analysis of LVEF measurement by Pearson’s correlation did not show statistically significant difference between different observers.
Conclusions: There was significant inter-observer variation with regard to the interpretation of valve abnormality, diastolic dysfunction and overall interpretation.
Key-words: Echocardiography, diastolic dysfunction, regional wall motion abnormality, valve abnormality.

INTRODUCTION
Echocardiography is one of the most widely used cardiac investigations. Although standard echocardiography is an accepted standard for the assessment of LV dysfunction, mitral regurgitation, aortic valve disease, and pericardial effusion, there remains substantial inter-observer variability in the interpretation of echocardiographic images [1-3]. Qualitative echocardiographic interpretation provides invaluable information for the practicing clinician. Nonetheless, quantitative echocardiography also plays a significant role in diagnosis, treatment, and prognostication. Given the cost and tremendous volume of echocardiograms performed every year, the interpretation of echocardiographic measurements would seem to be a critical area for quality control [4, 5]. With the emphasis on primary care and the explosion of imaging modalities, nonspecialist physicians and allied health care professionals are increasingly required to order and interpret the result of echocardiographic reports. It is virtually impossible for a nonspecialist to keep track of various reference limits [6, 7]. The interpretation of echocardiographic studies often involves the synthesis of single-dimensional quantitative measurements (obtained in select standard views) with qualitative assessment of 2-D images. As a final step, the composite echocardiographic information is correlated with clinical data available to the echocardiographer to generate a final report. The interpretation of echocardiographic measurements should consider the variability inherent in the imaging technique itself. Potential sources of variability in quantitative echocardiography include the patient, the sonographer, the techniques of image acquisition and analysis, the equipment, and the reader [8-11]. The present study is to evaluate the pattern of observer variability in the echocardiography interpretation.

SUBJECTS AND METHODS
A total of 37 patients were included in the study. Echocardiography was done in each patient four times at 2 weeks intervals by different cardiologists. Patients were examined thoroughly during each visit to confirm the stable disease. Echo reports were correlated with clinical evaluation of the patients. Each cardiologist was blinded to the previous reports. Echocardiography reports of each patient were analyzed for any discrepancy in interpretation. Parameters taken for evaluation were easily detectable, reproducible and were specific enough to detect discrepancies. These parameters are left ventricular ejection fraction, presence of diastolic dysfunction, presence of pulmonary hypertension, regional wall motion abnormality, valve abnormality and overall interpretation.
Statistical analysis: Inter-observer agreement was assessed by calculating the kappa ($\kappa$) value which was considered poor when the $\kappa$ value was <0.4, fair when the $\kappa$ value was between 0.4 and 0.59, good when the $\kappa$ value was between 0.6 and 0.79 and excellent when the $\kappa$ value was <0.8. Data analysis was performed using SPSS software. Variability in LVEF measurement was analysed by using Pearson’s correlation test. Statistical significance was assessed at a p value of less than 0.05.

RESULTS

Out of the 37 patients included in the study 20 were males and 17 were females. Mean age was 67+8.91 years. Analysis for agreement for detection of diastolic dysfunction in 4 different echo reports of the same patients done by different cardiologists showed fair agreement for reporter 1 and 2, echo1 and 3, reporter1 and 4, reporter2 and 3 (kappa statistics, table 1). Agreement was slight for reporter2 and 4, reporter3 and 4. Similarly for detection of valvular abnormality, there was only a slight agreement between reporter 1 and 2, reporter 2 and 3. Agreement between reporter1 and 3, reporter1 and 4, reporter2 and 4, reporter 3 and 4 were fair for detection of valvular abnormality (table 2). Regional wall motion abnormality discrepancy detected only in 3 reports and presence of pulmonary hypertension variability was seen in 2 patients. Discrepancy in the overall interpretation was seen in 14 patients. Agreement for overall interpretation of echocardiography showed better correlation for reporter 1 vs reporter3 and reporter2 vs reporter 3. Correlation for other pairs was slight (table 3).

Table 1: Inter-observer agreement for detection of diastolic dysfunction

<table>
<thead>
<tr>
<th>Reporter 1</th>
<th>Reporter 2</th>
<th>Reporter 3</th>
<th>Reporter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporter 1</td>
<td>0.407(0.73)</td>
<td>0.427(0.73)</td>
<td>0.370(0.679)</td>
</tr>
<tr>
<td>Reporter 2</td>
<td>0.313(0.676)</td>
<td>0.099(0.536)</td>
<td></td>
</tr>
<tr>
<td>Reporter 3</td>
<td>0.230(0.607)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as Kappa coefficient (Agreement).

Table 2: Inter-observer agreement for detection of Valvular abnormality

<table>
<thead>
<tr>
<th>Reporter 1</th>
<th>Reporter 2</th>
<th>Reporter 3</th>
<th>Reporter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporter 1</td>
<td>0.180(0.595)</td>
<td>0.460(0.757)</td>
<td>0.353(0.655)</td>
</tr>
<tr>
<td>Reporter 2</td>
<td>0.238(0.622)</td>
<td>0.322(0.655)</td>
<td></td>
</tr>
<tr>
<td>Reporter 3</td>
<td>0.470(0.724)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as Kappa coefficient (Agreement).

Table 3: Inter-observer agreement for overall echocardiography interpretation

<table>
<thead>
<tr>
<th>Reporter 1</th>
<th>Reporter 2</th>
<th>Reporter 3</th>
<th>Reporter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporter 1</td>
<td>0.2(0.778)</td>
<td>0.525(0.853)</td>
<td>0.195(0.742)</td>
</tr>
<tr>
<td>Reporter 2</td>
<td>0.464(0.829)</td>
<td>0.259(0.733)</td>
<td></td>
</tr>
<tr>
<td>Reporter 3</td>
<td>0.261(0.742)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as Kappa coefficient (Agreement).
Mean LVEF measured by reporter 1 was 61.45±3.22, reporter 2 was 61.09±2.45, reporter 3 was 60.65±4.81, and reporter 4 was 59.85±3.80. Analysis of LVEF measurement by Pearson's correlation did not show statistically significant difference between different observers (p>0.05, table 4). Difference observed between different observers for LVEF was clinically also not significant.

**DISCUSSION**

Echocardiographic laboratories make their measurements in several ways (M-mode, 2-D; on-line, off-line; with or without use of the American Society of Echocardiography guidelines)\[11, 12\]. Interobserver and intraobserver variability in measurements are well known even when measurements are made in a standard fashion. Furthermore, the stability of echocardiographic measurements is also influenced by the study quality, which in turn is affected by the age, weight, body habits of the patient, and presence of lung disease. Standardization of categories (i.e., mild, moderate, or severe) for describing the extent of deviation of observed measurements from reference limits is critical so that the descriptive terms used in echocardiography reports convey the same meaning to the nonspecialist who interprets and acts on them\[13\].

Present study showed considerable inter-observer variability in echocardiography interpretation. There was significant inter-observer variation with regard to the detection of diastolic dysfunction and valve abnormality ($r$ values between 0.09-0.42 and 0.18 to 0.47 respectively). Also the agreement for overall interpretation was slight to moderate ($k$ values between 0.2-0.5). There was no statistically significant variation in LVEF measurements between different observers (p>0.05). As we have not recognized any other causative factors, variations which were observed considered as purely as inter-individual variability. In a study across 3 institutions and involving nearly a dozen experienced operators, only a quarter of images obtained were adjudged to be optimally positioned according to standard 2-D imaging views. Thus there is "subjectivity" in the interpretation of "standardized" imaging planes and "standard" transducer positions that explains variability in image acquisition\[14\].

Limitations of the present study should be considered. Echocardiography was done at different time points i.e. at interval of 2 weeks. Though the stable condition of the patients were confirmed clinically, temporal variation occurring in the underlying condition cannot be completely ruled out. This factor also may be contributing for the observed disagreement in
interpretation of results of different echocardiography of the same patients. Ideally, same image should be read by different observers in a blinded fashion.

Important aspect of quality which is required for accreditation of any laboratories is production of consistent, accurate and reproducible data. Each physician should be consistent. Re reading the same echo should generate equivalent reports. Intermittent testing of variability will assure physician consistency. Referring physician should be confident that the every reading physician will interpret a patient in a consistent reproducible manner. Intra and inter-observer variability studies help to detect interpretation problem, documents physician inadequacies, allows problem to be rectified. Echo findings should be verified by comparing the results with other diagnostic modalities that generate similar results.

CONCLUSION

Present study has shown considerable variability in echocardiography interpretation especially with regard to presence of valvular abnormality, diastolic dysfunction and overall interpretation. Variability in the assessment of LVEF was clinically as well as statistically not significant and hence may be considered as a reliable parameter.
REFERENCES


