Development and Validation of a Teaching Module for Echocardiographic Scoring of Rheumatic Mitral Stenosis

Nathan Stehouwer*,†, Emmy Okello‡, Vedant Gupta§, Alison L. Bailey§, Richard Josephson*,†, Sri Krishna Madan Mohan*,†, Mohammed N. Osman*,†, Chris T. Longenecker*,†

Cleveland, OH, USA; Kampala, Uganda; and Lexington, KY, USA

ABSTRACT

Background: The Wilkins score and commissural calcification scores predict outcomes after percutaneous balloon mitral valvuloplasty. However, many cardiologists are inadequately trained in their application—both in the United States where the incidence of rheumatic heart disease has fallen and in rheumatic heart disease endemic countries where training infrastructure is weak.

Objectives: This study sought to develop a computer-based educational module teaching 2 scoring systems for rheumatic mitral stenosis and to validate the module among cardiology fellows in the United States and Uganda.

Methods: We developed a module organized into 3 sets of 10 echocardiograms each. The module was completed by 13 cardiology fellows from 2 academic centers in the United States and 1 in Uganda. Subject answers were compared with a score assigned by 2 experts in echocardiography. The primary outcome was change in subjects' accuracy from set 1 to set 3, measured by mean absolute deviation from expert scores. Secondary outcomes included change in interoperator variability and individual subject bias from set 1 to set 3.

Results: The mean absolute deviations from expert scores in sets 1 and 3 were 2.09 and 1.82 for the Wilkins score (possible score range 0 to 16) and 1.13 and 0.94 for the commissural calcification score (possible score range 0 to 4). The change from set 1 to set 3 was statistically significant only for 1 of the Wilkins component scores (leaflet calcification, $p < 0.001$.) No change was seen in the interoperator variability. Individual subject bias in assigning the total Wilkins score was reduced from set 1 to set 3.

Conclusions: Use of this module has the potential to enhance the training of cardiologists in the echocardiographic assessment of mitral stenosis. Modified versions of this module or similar ones should be tested in targeted populations of cardiology trainees with the most exposure to mitral stenosis interventions.

C.T. Longenecker is supported by the National Institutes of Health (grant K23 HL123341), a Wolf Family Foundation Scholars Grant, and Medtronic Philanthropy; consults for Gilead Sciences; and has received research grants from Bristol-Myers Squibb and Medtronic Philanthropy. All other authors report no relationships that could be construed as a conflict of interest.
This work was supported in part by Medtronic Philanthropy.
The sponsors had no role in study design, data collection, analysis, interpretation of results, the writing of this manuscript, or in the decision to submit for publication.

From the *Case Western Reserve University School of Medicine, Cleveland, OH, USA; †University Hospitals Cleveland Medical Center, Cleveland, OH, USA; ‡Uganda Heart Institute, Kampala, Uganda; and the §Gill Heart Institute, University of Kentucky, Lexington, KY, USA. Correspondence: C. T. Longenecker (chris.longenecker@uhhospitals.org).

GLOBAL HEART

VOL. 13, NO. 2, 2018
ISSN 2211-8160/$36.00.
http://dx.doi.org/10.1016/j.gheart.2017.05.004

Wilkins et al. [1] reported a novel scoring system in 1988 to predict success of percutaneous balloon mitral valvuloplasty (PBMV) for the treatment of mitral stenosis. The Wilkins scoring system relies on 4 echocardiographic characteristics of the mitral valve: leaflet mobility, leaflet thickness, leaflet calcification, and subvalvular thickening. Each variable is scored on a scale of 0 to 4 with higher scores representing more severe involvement. The additive score of 0 to 16 was found to predict outcomes, with higher scores (>8) predictive of suboptimal outcomes.

Several alternative echocardiographic systems for prediction of post-PBMV outcomes have been developed subsequently, including commissural calcification (CC) scoring [2,3]. The CC system, which our module teaches and was published by Sutaria et al. [2] in 2000, scores each quadrant of the commissure 0 or 1 based on the presence of calcification for a total score 0 to 4, with lower scores predicting better outcomes after PBMV. Other proposed scoring systems for pre-procedural assessment of mitral stenosis include assessment of mitral annular calcification [4]; ratio of posterior to anterior mitral valve leaflets [5]; a novel score proposed by Iung and Cormier [6] based on valve flexibility, subvalvular fusion, and leaflet calcification; and a scoring system using real-time 3-dimensional echocardiography [7]. Of these alternative scores, scoring of commissural morphology has shown particularly strong evidence for prediction of post-PBMV results [8].

Case-based modules have been successfully used to teach skills in other areas of echocardiography, including assessment of left ventricular function [9] and a 3-dimensional echocardiography module to enhance assessment of transesophageal echocardiography [10].

TABLE 1. Expert scores of each mitral stenosis case included in the module

Set	Case	Wilkins Score: Wilkins Total (0–16)	Wilkins Score: Leaflet Mobility (0–4)	Wilkins Score: Leaflet Thickness (0–4)	Wilkins Score: Leaflet Calcification (0–4)	Wilkins Score: Subvalvular Thickening (0–4)	Commissural Calcification (0–4)
Set 1	Case 1	9	2	2	2	3	1
	Case 2	12	3	3	3	3	1
	Case 3	4	1	1	1	1	3
	Case 4	13	3	4	3	3	2
	Case 5	7	1	2	2	2	3
	Case 6	11	2	3	4	2	1
	Case 7	14	4	3	4	3	1
	Case 8	7	2	2	1	2	1
	Case 9	7	1	1	3	2	1
	Case 10	6	2	2	0	2	1
Set 1 median (Q1, Q3)		8 (7, 12)	2.5 (1, 3)	2 (2, 3)	2.5 (1, 3)	2 (2, 3)	1 (1, 2)
Set 2	Case 11	5	1	2	1	1	1
	Case 12	8	2	2	2	2	0
	Case 13	5	1	1	1	2	0
	Case 14	10	3	2	3	2	1
	Case 15	5	1	1	1	2	
	Case 16	8	2	2	2	2	2
	Case 17	9	2	2	2	3	1
	Case 18	14	4	4	4	2	4
	Case 19	6	1	2	1	2	2
	Case 20	7	2	2	1	2	0
Set 2 median (Q1, Q3)		7.5 (5, 9)	2 (1, 2)	2 (2, 2)	1.5 (1, 2)	2 (2, 2)	1 (0, 2)
Set 3	Case 21	10	3	2	3	2	2
	Case 22	11	3	3	3	2	1
	Case 23	6	1	2	1	2	1
	Case 24	4	1	1	1	1	1
	Case 25	9	2	3	2	2	1
	Case 26	8	2	2	2	2	1
	Case 27	6	1	2	1	2	0
	Case 28	8	2	2	2	2	1
	Case 29	8	2	2	2	2	2
	Case 30	4	1	1	1	1	1
Set 3 median (Q1, Q3)		8 (6, 9)	2 (1, 2)	2 (2, 2)	2 (1, 2)	2 (2, 2)	1 (1, 1)
Module Median (Q1, Q3)		8 (6, 10)	2 (1, 2)	2 (2, 2)	2 (1, 3)	2 (2, 2)	1 (1, 2)
p Value	Set 1 vs. set 2	0.28	0.56	0.43	0.34	0.28	0.85
	Set 1 vs. set 3	0.24	0.50	0.47	0.30	0.10	0.27
	Set 2 vs. set 3	0.82	0.82	1.00	1.00	0.34	0.81

Q, quartile.

Our project had 2 objectives. The first objective was to develop a computer-based module that teaches correct application of the Wilkins and CC scoring schema of the mitral valve. The second was to validate the module's use by cardiology fellows in Uganda and the United States. We hypothesized that use of this teaching module by cardiology fellows would improve accuracy, reduce interoperator variability (IOV), and reduce individual subject bias in the application of the Wilkins and CC scores for mitral stenosis.

METHODS

Development of the teaching module

Thirty representative transthoracic echocardiograms were selected, 20 from the University Hospitals-Harrington Heart and Vascular Institute (HHVI) database in Cleveland, OH, USA, and 10 from the Uganda Heart Institute in Kampala, Uganda. Echocardiograms were selected for presence of rheumatic mitral stenosis, transvalvular gradient >5 mm Hg, valve area <1.5 cm^2, and no more

TABLE 2. Inter-reader agreement of experts by set

	Wilkins Score					
	Total Wilkins Score	Leaflet Mobility	Leaflet Thickness	Valve Calcification	Subvalvular Thickening	Commissural Calcification
Set 1	0.830 (0.444–0.958)	0.900 (0.632–0.976)	0.724 (0.207–0.929)	0.890 (0.598–0.974)	0.231 (−0.380 to 0.740)	0.033 (−0.165 to 0.449)
Set 2	0.793 (0.361–0.948)	0.869 (0.514–0.969)	0.529 (−0.120 to 0.868)	0.644 (−0.028 to 0.909)	0.355 (−0.260 to 0.796)	0.422 (−0.127 to 0.826)
Set 3	0.825 (0.397–0.962)	0.907 (0.637–0.980)	0.621 (−0.147 to 0.913)	0.755 (0.238–0.944)	−0.324 (−0.997 to 0.506)	−0.135 (−0.254 to 0.386)
Overall	0.822 (0.642–0.916)	0.886 (0.764–0.947)	0.631 (0.324–0.816)	0.821 (0.645–0.915)	0.192 (−0.544 to 0.699)	0.157 (−0.099 to 0.457)

Values are intraclass correlations (95% confidence intervals).

than moderate mitral regurgitation. Echocardiograms were selected to represent a range of pathologic severity.

Mitral valve scores were determined by the average of the assessments of 2 expert readers working independently, with significant disagreements (score difference of >2) adjudicated by a third expert. The experts are HHVI attending echocardiographers who have extensive experience both in the United States and in areas of higher rheumatic heart disease (RHD) prevalence, and both have evaluated at least 200 cases of rheumatic mitral stenosis.

The module was presented as a slideshow with embedded videos of the selected views. Cases were organized into 3 sets of 10 cases each. We began by dividing cases at random into the 3 sets and then modified the distribution as necessary to ensure that each set had a comparable average Wilkins and CC score and comparable image quality. Subjects were shown a brief explanation of the Wilkins and CC scores and were then immediately asked to apply the scoring system to the first set of 10 cases as a pre-test. The subjects then proceeded to the didactic portion of the module, which begins with several slides with information and discussion on the application of Wilkins and CC scoring, followed by example slides for each of the Wilkins components and the CC score. Each slide shows several examples drawn from the pre-test, accompanied by an explanation of the correct scoring. Following these examples, the subjects scored an additional 10 cases (set 2). During these cases, subjects were shown the expert score immediately following each case. Finally, subjects scored the final set of 10 cases as a post-test. Subjects did not have access to the expert scores for the post-test. The module was designed so that it could be completed by a fellow in 1 sitting over a 2- to 3-h period, but subjects were permitted to complete the module at their own pace.

Subject selection and recruitment

Subjects to validate the teaching cases were recruited on a voluntary basis from fellows in the HHVI, Uganda Heart Institute, and University of Kentucky Gill Heart Institute cardiology fellowship programs. There was no requirement for a particular amount of prior echocardiography experience. Participation and results were collected with complete anonymity.

Outcomes

The primary outcome was 1) the change in accuracy—defined as the mean absolute value of deviation from expert scores on individual cases—for the total Wilkins score from set 1 to set 3. Secondary outcomes were 2) change in accuracy for each of the Wilkins component scores separately from set 1 to set 3; 3) change in IOV of the Wilkins and CC scores from set 1 to set 3; and 4) change in IOV of the Wilkins component scores from set 1 to set 3. We also analyzed the change in subject bias from set 1 to set 3, measured as the mean deviation for each subject by set.

Statistical analysis

Accuracy was assessed by measuring the mean of the absolute value of the deviation of a subject answer from the expert answer on each individual case. The average of these absolute deviations for each set (10 cases with 13 subject answers each) was then reported as the mean absolute deviation. Significance in changes in mean absolute deviation were compared using the Wilcoxon test for paired observations. Intraclass correlation was used to assess IOV among the expert readers and IOV of the cardiology fellows' scores separately by case set. Subject bias was assessed for each subject by calculating the mean of the deviation from expert scores for each of the 10 cases in 1 set. Analyses were performed using GraphPad Prism for Mac OS X (version 6.0h; GraphPad Software, La Jolla, CA, USA) and STATA (version 14.0; StataCorp, College Station, TX, USA). A p value of <0.05 was considered statistically significant.

RESULTS

The average expert scores for each individual case used in the teaching module are displayed in Table 1. The severity of cases was well balanced among the 3 sets without significant differences among the sets in terms of mean Wilkins score, CC score, or any of the Wilkins components (all $p > 0.10$). Agreement between the 2 expert scorers is shown in Table 2. High intraclass correlation was seen between experts for the total Wilkins score and leaflet characteristics of mobility, thickness, and calcification. Low intraclass correlation was seen for subvalvular thickening, which is consistent with

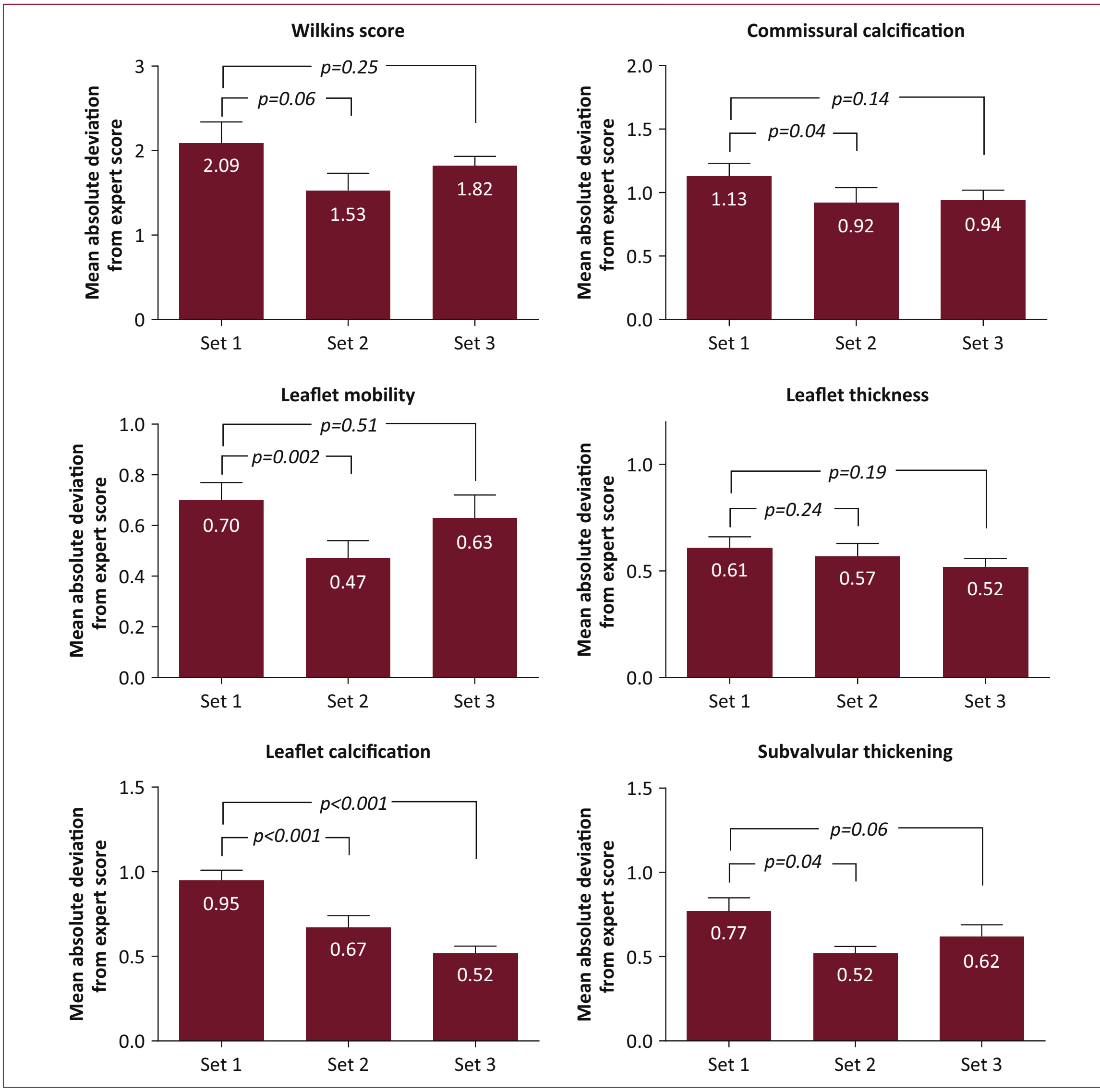

FIGURE 1. Mean absolute deviation of trainee score compared with expert score.

previous reports [11], and commissural calcification scoring.

Thirteen subjects completed the teaching module during the study period, 7 from the United States and 6 from Uganda. There was a decrease in the absolute deviation from the expert score from set 1 to set 3 for the Wilkins score, the CC score, and each of the 4 components of the Wilkins score (Figure 1) though this change was statistically significant only for leaflet calcification ($p < 0.001$). IOV between subjects, measured by intraclass correlation (Table 3), did not change appreciably from set 1 to set 3.

Individual subject bias was assessed by the average deviation from expert score by set (Figure 2). For overall Wilkins score, the mean subject bias was lower in set 3 than in set 1 (1.23 vs. 0.62, set 1 vs. 3; $p = 0.01$).

DISCUSSION

Although many prior studies report the predictive value and IOV of scoring systems for mitral stenosis (Table 4) [12], there has been relatively little prior study of how readily these scoring systems can be taught to nonexperts. Our study is the first to report on use of computer-based simulation to teach the Wilkins and CC scoring systems. In our study, use of a computer-based teaching module was associated with a trend toward improved accuracy in scoring of mitral stenosis, though the improvement was statistically significant in only 1 of the component scores.

TABLE 3. Intersubject variability by set

	Wilkins Score					
	Total Wilkins Score	Leaflet Mobility	Leaflet Thickness	Leaflet Calcification	Subvalvular Thickening	Commissural Calcium
Set 1	0.458 (0.223–0.815)	0.354 (0.148–0.747)	0.406 (0.204–0.732)	0.345 (0.159–0.680)	0.252 (0.098–0.590)	0.110 (0.024–0.357)
Set 2	0.608 (0.395–0.845)	0.601 (0.388–0.841)	0.427 (0.227–0.730)	0.549 (0.336–0.812)	0.246 (0.093–0.565)	0.328 (0.135–0.698)
Set 3	0.438 (0.237–0.738)	0.302 (0.135–0.621)	0.369 (0.181–0.684)	0.471 (0.265–0.762)	0.109 (0.017–0.367)	0.032 (−0.034 to 0.280)

Values are intraclass correlations (95% confidence intervals).

We did not see a change in intersubject variability in scoring from set 1 to set 3. For subvalvular thickening and CC score, there was poor inter-reader agreement among both experts and trainees. Individual subject bias was reduced from set 1 to set 3.

The incidence of RHD has declined significantly in North America [13]. As a result, cardiologists in training in the United States receive relatively little experience evaluating mitral stenosis. In contrast, cardiology trainees in Sub-Saharan Africa have frequent exposure to RHD but

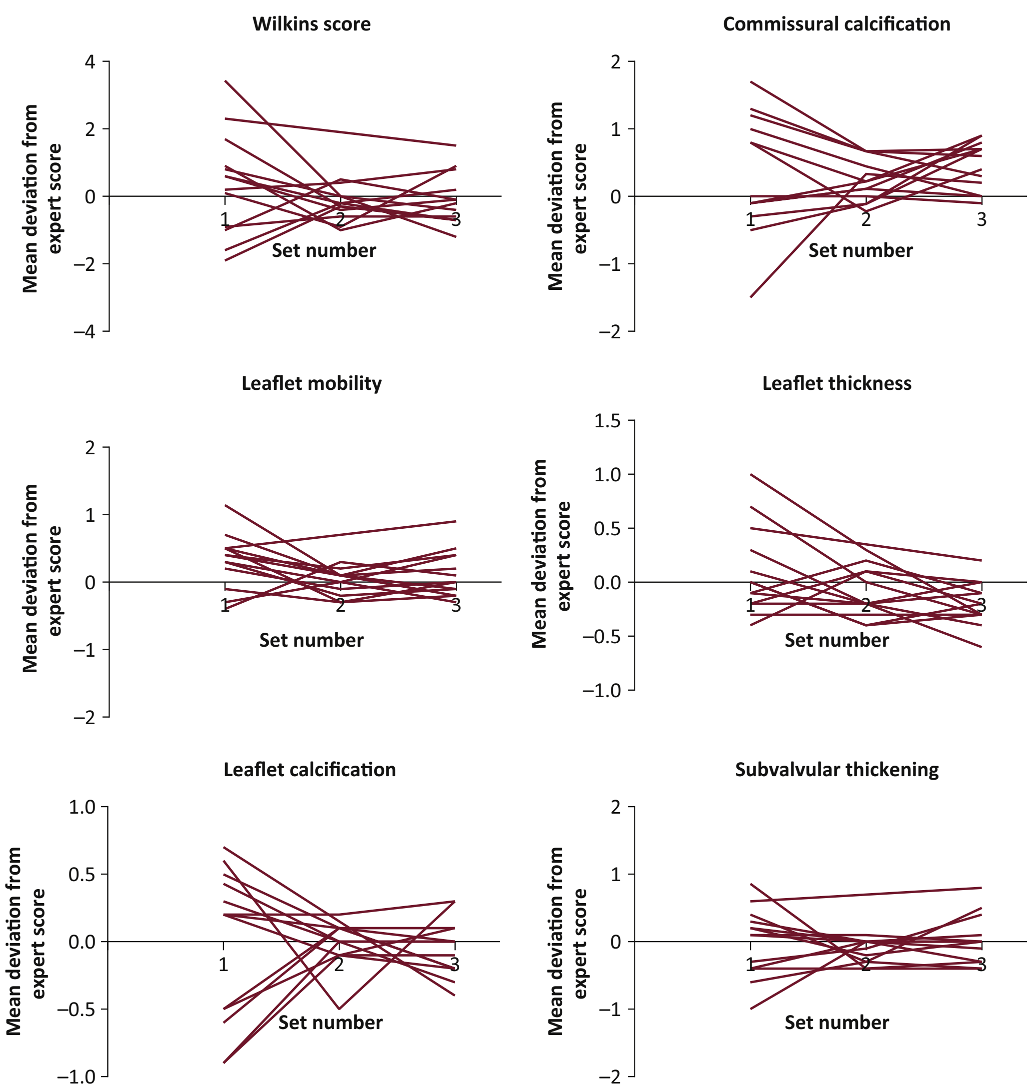

FIGURE 2. Mean difference from expert score by individual subject.

TABLE 4. Reported intraobserver and interobserver variability of expert application of Wilkins score

	Intraobserver	Interobserver	Unit of Measurement
Wilkins et al. [1]	0.41	0.38	Mean unsigned difference
Salarifar et al. [4]	7.50%	11.10%	% Difference
Components	0.58—0.68	0.48—0.62	Kappa
Anwar et al. [11]			
Thickness	0.65	0.55	Kappa
Mobility	0.58	0.56	Kappa
Calcification	0.41	0.03	Kappa
Subvalvular involvement	0.14	0.01	Kappa
Veyrat et al. [12]	3.75%	4.73%	% Difference

are without access to PBMV in many locations and may not have been exposed to the use of these scoring systems. The Wilkins and commissural calcification scores, being based on subjective evaluation, could vary in application between subject experts and nonexperts. Thus far, intra- and interobserver variability of the Wilkins score have only been reported for expert readers.

This study highlights the difficulty of developing a gold standard for teaching a subjective scoring system. This difficulty was highlighted by the low intraclass correlation between experts seen for subvalvular thickening and commissural calcification. Of note, the low expert agreement in assessing subvalvular thickening has been reported by other groups as well [11]. Additionally, this study may not have been adequately powered to detect a difference between set 1 and set 3, since a consistent trend toward improvement was seen that did not meet statistical significance for most component scores. It is also possible that exposure to a greater number of cases would have resulted in a continued downward trend in deviation from expert scores. Our study was not designed to assess the number of cases required for competence in these scoring systems, and this would be an important area for further research.

Our study has several strengths. The use of echocardiograms from patients both in the United States and in Sub-Saharan Africa increases the generalizability of the content taught by the module. Validation by trainees in both the United States and Uganda also adds to the generalizability of the study findings. By reporting results for all component scores, our study demonstrates that some of the components are highly variable and thus more difficult to teach.

This is the first study to assess an educational module designed to teach the Wilkins and CC scores to fellows in training. Among cardiologists practicing only in Western countries, this skill set may be most important for those who specialize in echocardiography or interventional cardiology, but it is important to demonstrate the feasibility of teaching echocardiographic scoring systems to a general cardiology fellow for those fellows who plan to practice in, or are training in, a resource-limited setting, as well as for fellows who plan to specialize in imaging or intervention.

Simulation is an increasingly discussed and implemented tool in cardiology training [14,15]. Use of simulation may be an important tool in medical education as part of capacity-building projects in resource-limited settings. Our teaching module could be used as a simulation for development and maintenance of competence in this area in cardiology trainees in the United States and in countries with higher prevalence of RHD. In Uganda, the first PBMV was performed at the Uganda Heart Institute in 2012 by a team from the HHVI. Since that time, 2 Ugandan interventional cardiologists have been trained in Cleveland and since December 2016 are now *independently* performing percutaneous mitral valve interventions. Future studies in resource-limited settings should evaluate the usefulness of the Wilkins and CC scores to predict outcomes and should assess the utility of refined teaching modules among a growing number of cardiology trainees in these countries.

CONCLUSIONS

Use of a computer-based module consisting of 30 cases of rheumatic mitral stenosis tended to improve scoring of valve suitability for PBMV in mitral stenosis and decreased subject bias, but there was significant heterogeneity among subcomponents of the Wilkins score. Leaflet calcification improved the most, but subvalvular thickening and commissural calcification scores showed substantial interobserver variability among both experts and trainees. This educational module has the potential to enhance fellow training in echocardiographic assessment of the mitral valve and mitral stenosis. Modified versions of this module or similar ones should be tested in targeted populations of cardiology trainees with the most exposure to mitral stenosis interventions.

REFERENCES

1. Wilkins GT, Weyman AE, Abascal VM, Block PC, Palacios IF. Percutaneous balloon dilatation of the mitral valve: an analysis of echocardiographic variables related to outcome and the mechanism of dilatation. Br Heart J 1988;60:299–308.
2. Sutaria N, Northridge DB, Shaw TRD. Significance of commissural calcification on outcome of mitral balloon valvotomy. Heart 2000;8: 398–402.

3. Padial LR, Freitas N, Sagie A, Newell JB, Weyman AE, Levine RA, Palacios IF. Echocardiography can predict which patients will develop severe mitral regurgitation after percutaneous mitral valvulotomy. J Am Coll Cardiol 1996;27:1225–31.
4. Salarifar M, Rezvanfard M, Sadeghian H, Safir-Mardanloo A, Shafii N. Mitral annular calcification predicts immediate results of percutaneous transvenous mitral commissurotomy. Cardiovasc Ultrasound 2011;9:29.
5. Mahfouz RA. Utility of the posterior to anterior mitral valve leaflets length ratio in prediction of outcome of percutaneous balloon mitral valvuloplasty. Echocardiography 2011;28:1068–73.
6. Iung B, Cormier B, Porte J, Nallet O, Michel P, Acar J, Vahanian A. Immediate results of percutaneous mitral commissurotomy: a predictive model on a series of 1514 patients. Circulation 1996;94:2124–30.
7. Soliman OII, Anwar AM, Metawei AK, McGhie JS, Geleijnse ML, Ten Cate FJ. New scores for the assessment of mitral stenosis using real-time three-dimensional echocardiography. Curr Cardiovasc Imaging Rep 2011;4:370–7.
8. Fatkin D, Roy P, Morgan JJ, Feneley MP. Percutaneous balloon mitral valvotomy with the Inoue single-balloon catheter: commissural morphology as a determinant of outcome. J Am Coll Cardiol 1993;21:390–7.
9. Johri AM, Picard MH, Newell J, Marshall JE, King MEE, Hung J. Can a teaching intervention reduce interobserver variability in LVEF assessment. J Am Coll Cardiol Img 2011;4:821–9.
10. Jerath A, Vegas A, Meineri M, Silversides C, Feindel C, Beattie S. An interactive online 3D model of the heart assists in learning standard transesophageal echocardiography views. Can J Anesth 2011;58:14–21.
11. Anwar AM, Attia WM, Nosir YF, Soliman OI, Mosad MA, Othman M, et al. Validation of a new score for the assessment of mitral stenosis using real-time three-dimensional echocardiography. J Am Soc Echocardiogr 2010;23:13–22.
12. Veyrat C, Pellerin D, Sainte Beuve D, Larrazet F, Kalmanson D, Witchitz S. Colour Doppler valvar and subvalvar flow diameter imaging versus echo score in mitral stenosis: comparison with type of surgery. Heart 1996;75:486–91.
13. Tibazarwa KB, Volmink JA, Mayosi BM. Incidence of acute rheumatic fever in the world: a systematic review of population-based studies. Heart 2008;94:1534–40.
14. Fox K. Simulation-based learning in cardiovascular medicine: benefits for the trainee, the trained and the patient. Heart 2012;98:527–8.
15. De Ponti R, Marazzi R, Ghiringhelli S, Salerno-Uriarte JA, Calkins H, Cheng A. Superiority of simulator-based training compared with conventional training methodologies in the performance of transseptal catheterization. J Am Coll Cardiol 2011;58:359–63.