Published in final edited form as:

Glob Heart. 2018 June ; 13(2): 105–111. doi:10.1016/j.gheart.2017.05.004.

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

Nathan Stehouwer, MD^{a,b}, Emmy Okello, M.Med, PhD^c, Vedant Gupta, MD^d, Alison L. Bailey, MD^{d,1}, Richard Josephson, MS, MD^{a,b}, Sri Krishna Madan Mohan, MD^{a,b}, Mohammed Osman, MD^{a,b}, and Chris T. Longenecker, MD^{a,b}

^aCase Western Reserve University School of Medicine, 2109 Adelbert Rd, Cleveland, OH 44106

^bUniversity Hospitals Cleveland Medical Center, 11100 Euclid Ave, Cleveland, OH ^cUganda Heart Institute, P.O. Box 7051, Kampala Binaisa Road, Kampala, Uganda ^dGill Heart Institute, 900 S Limestone, University of Kentucky, Lexington, KY, 40508

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 (RHD) has fallen but also in RHD endemic countries where training infrastructure is weak.

Objectives—The objectives were (1) to develop a computer based educational module teaching two scoring systems for rheumatic mitral stenosis, and (2) to validate the module among cardiology fellows in the United States and Uganda.

Study Design-Setting-Patients—We developed a module organized into 3 sets of 10 echocardiograms each. The module was completed by 13 cardiology fellows from two academic centers in the United States and one in Uganda. Subject answers were compared with a score assigned by two experts in echocardiography.

Outcome Measures—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.

Findings and Interpretation—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 0–16) and 1.13 and 0.94 for the commissural calcification score (possible score range 0–4). The change from set 1 to set 3 was statistically significant only for one of the Wilkins component scores (leaflet calcification,

Corresponding Author: Chris T. Longenecker, MD.

¹Present address: Erlanger Health System/University of Tennessee COM Chattanooga, 960 E 3rd St, Chattanooga, TN 37403

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosures:

No other authors had any disclosures.

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. 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.

Keywords

Wilkins; Rheumatic Mitral Stenosis; valvuloplasty; education; simulation

Introduction

Wilkins et al reported a novel scoring system in 1988 to predict success of percutaneous balloon mitral valvuloplasty (PBMV) for the treatment of mitral stenosis¹. 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 0–4 with higher scores representing more severe involvement. The additive score of 0–16 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, published by Sutaria et al in 2000, scores each quadrant of the commissure 0–1 based upon the presence of calcification for a total score 0–4, with lower scores predicting better outcomes after PBMV². Other proposed scoring systems for pre-procedural assessment of mitral stenosis include assessment of mitral annular calcification⁴, ratio of posterior to anterior mitral valve leaflets⁵, a novel score proposed by Lung and Cormier based on valve flexibility, subvalvular fusion, and leaflet calcification⁶, and a scoring system using real time 3D echocardiography⁷. Of these alternative scores, scoring of commissural morphology has shown particularly strong evidence for prediction of post-PBMV results⁸.

Case-based modules have been successfully used to teach skills in other areas of echocardiography, including assessment of left ventricular function⁹ and a 3D echocardiography module to enhance assessment of transesophageal echocardiography¹⁰. Our project had two 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.

Material and Methods

Development of the teaching module

Thirty representative transthoracic echocardiograms were selected, twenty from the UH-Harrington Heart & Vascular Institute (HHVI) database in Cleveland, OH and ten from the Uganda Heart Institute (UHI) in Kampala, Uganda. Echocardiograms were selected for presence of rheumatic mitral stenosis, transvalvular gradient $>5\text{mmHg}$, valve area $<1.5\text{cm}^2$, and no more 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 two 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 three 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 one sitting over a 2–3 hour 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, UHI and University of Kentucky Gill Heart Institute (GHI) 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 interoperator variability (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 one 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 <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 between the sets in terms of mean Wilkins score, CC score, or any of the Wilkins components (all $p>0.10$). Agreement between the two expert scorers is shown in Table 3. 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, consistent with previous reports¹¹, and commissural calcification scoring.

Thirteen subjects completed the teaching module during the study period, seven from the United States and six 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 four 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 4), 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

While many prior studies report the predictive value and IOV of scoring systems for mitral stenosis (Table 4), there has been relatively little prior study of how readily these scoring systems can be taught to non-experts. 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 one

of the component scores. We did not see a change in inter-subject 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 America¹². 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 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 non-experts. Thus far, intra- and inter-observer 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 well¹¹. 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 which 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 commissural calcification 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^{13,14}. 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

percutaneous balloon mitral valvuloplasty was performed at the Uganda Heart Institute in 2012 by a team from the HHVI. Since that time, two 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 sub-components of the Wilkins score. Leaflet calcification improved the most, but subvalvular thickening and commissural calcification scores showed substantial inter-observer 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.

Acknowledgments

This work was supported in part by Medtronic Philanthropy, and by National Institutes of Health [grant number K23 HL123341 to CTL]. 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.

CTL is a consultant for Gilead Sciences and has received research grants from Bristol-Myers Squibb and Medtronic Philanthropy.

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(4):299–308. [PubMed: 3190958]
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, et al. Echocardiography Can Predict Which Patients Will Develop Severe Mitral Regurgitation after Percutaneous Mitral Valvulotomy. *J Am Coll Cardiol*. 1996; 27(5):1225–31. [PubMed: 8609347]
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(1):29. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3217870&tool=pmcentrez&rendertype=abstract>. [PubMed: 22035075]
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(10):1068–73. [PubMed: 21966895]
6. Iung B, Cormier B, Porte J, Nallet O, Michel P, Acar J, et al. Immediate Results of Percutaneous Mitral Commissurotomy. A Predictive Model on a Series of 1514 Patients. *Circulation*. 1996; 94(9): 2124–30. [PubMed: 8901662]
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(5):370–7. [PubMed: 21949566]

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(2):390–7. [PubMed: 8426003]
9. Johri AM, Picard MH, Newell J, Marshall JE, King MEE, Hung J. Can a Teaching Intervention Reduce Interobserver Variability in LVEF Assessment. *JACC Cardiovasc Imagin.* 2011; 4(8):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(1):14–21. [PubMed: 21069586]
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(1):13–22. [PubMed: 19926444]
12. Tibazarwa KB, Volmink JA, Mayosi BM. Incidence of acute rheumatic fever in the world : a systematic review of population-based studies. *Heart.* 2008; 94(12):1534–40. [PubMed: 18669552]
13. Fox K. Simulation-based learning in cardiovascular medicine: benefits for the trainee, the trained and the patient. *Heart.* 2012; 98(7):527–8. [PubMed: 22337950]
14. 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(4):359–63. [PubMed: 21757112]
15. 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(5):486–91. [PubMed: 8665342]

Highlights

- Use of a teaching module for echocardiographic scoring of mitral stenosis was associated with a non-significant improvement in subject accuracy from pre-test to post-test in overall Wilkins and Commissural Calcification scores.
- Subject accuracy in assessment of leaflet calcification improved from pre-test to post-test ($p < 0.001$.)
- Intersubject variability in the application of the Wilkins and Commissural Calcification scores did not substantially differ from pre-test to post-test.
- For overall Wilkins score, the mean subject bias was lower in the post-test than the pretest ($p < 0.05$).

Figure 1.
Mean absolute deviation of trainee score compared to expert score

Figure 2.
Mean difference from expert score by individual subject

Table 1

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

Set	Case	Wilkins Score				Commissural Calcification (0-4)	
		Wilkins Total (0-16)	Leaflet Mobility (0-4)	Leaflet Thickness (0-4)	Leaflet Calcification (0-4)	Subvalvular Thickening (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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Set	Case	Wilkins Total (0–16)	Leaflet Mobility (0–4)	Leaflet Thickness (0–4)	Leaflet Calcification (0–4)	Subvalvular Thickening (0–4)	Commissural Calcification (0–4)
	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

Table 2

Inter-reader agreement of experts by set

Results shown as intraclass correlation (95% confidence interval)

	Wilkins Score			Commissural calcification		
	Total Wilkins Score	Leaflet Mobility	Leaflet Thickness	Valve calcification	Subvalvular Thickening	
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 (–0.380–0.740)	0.033 (–0.165–0.449)
Set 2	0.793 (0.361–0.948)	0.869 (0.514–0.969)	0.529 (–0.120–0.868)	0.644 (–0.028–0.909)	0.355 (–0.260–0.796)	0.422 (–0.127–0.826)
Set 3	0.825 (0.397–0.962)	0.907 (0.637–0.980)	0.621 (–0.147–0.913)	0.755 (0.238–0.944)	–0.324 (–0.997–0.506)	–0.135 (–0.254–0.386)
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 (–0.544–0.699)	0.157 (–0.099–0.457)

Table 3

Inter-subject variability by set

Shown as intraclass correlation (95% confidence interval)

	Wilkins Score				Commissural calcium	
	Total Wilkins Score	Leaflet Mobility	Leaflet Thickness	Leaflet Calcification	Subvalvular Thickening	
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 (–0.034–0.280)

Table 4
Reported intra-observer and inter-observer variability of expert application of Wilkins score

Sources: Wilkins et al 1988¹, Salarifar et al 2011⁴, Anwar et al 2010¹¹, Veyrat et al 1996¹⁵

	Intra-observer	Inter-observer	Unit of measurement
Wilkins	0.41	0.38	mean unsigned difference
Salarifar	7.50%	11.10%	% difference
<i>Components</i>	0.58–0.68	0.48–0.62	kappa
Anwar			
<i>Thickness</i>	0.65	0.55	kappa
<i>Mobility</i>	0.58	0.56	kappa
<i>Calcification</i>	0.41	0.03	kappa
<i>Subvalvar Involvement</i>	0.14	0.01	kappa
Veyrat	3.75%	4.73%	% difference