Impact of variability in echocardiographic interpretation on assessment of repair following congenital heart surgery

A. Parthiban[1] and J. C. Levine[2], M. Nathan[2], J. A. Marshal[1], G. S. Shirali[1], S. D. Simon[1], S. D. Colan[2], J. W. Newburger[3], G. Raghuveer[1]

[1]Children's Mercy Hospital, University of Missouri Kansas City School of Medicine, [2]Boston Children's Hospital, [3]Harvard University

Background: Technical Performance Score (TPS), a tool based largely on the presence and magnitude of residua on postoperative echocardiograms (echo), has been used for assessing surgical repair and correlates with outcomes. The reproducibility of the echo measures that drive TPS classification has not been tested. We evaluated reader variability for echo components of TPS for tetralogy of Fallot (TOF) repair and arterial switch operation (ASO) in 2 centers and measured its effect on TPS.

Hypothesis: Inter-reader echo measurement variability will not substantially impact TPS classification.

Methods: Postoperative echos were evaluated in 67 subjects (39 TOF and 28 ASO). Two readers (1 per center) read each echo, blinded to center of origin. To assess intra-reader variability, 25% of echoes were re-reviewed by each reader. Measurements between readers were compared with Intra-class correlation (ICC). TPS Class (1 Optimal *no residua*, 2 Adequate *minor residua*, 3 Inadequate *major residua*) was assigned for each echo review by an independent investigator. The impact of measurement variability on overall TPS variability was compared using weighted Kappa (K) and % raw agreement.

Results: ICC was highest for Doppler velocity data and lower for measurements of small linear structures such as septal defects and vena contracta *Figure.* Overall TPS demonstrated good agreement (between reader TOF K = 0.82 and ASO K = 0.81). The 2 readers were concordant for TPS Class for 53 subjects (79%) and discordant for Classes 2 vs. 3 in 6 (9%); no readings were discordant between Classes 1 and 3 *Table*.

Conclusions: Although overall TPS demonstrated good agreement, inter-reader variation for echo measurements had a small, but important effect on TPS for ASO and TOF, particularly for the distinction between minor and major residua. Future studies of generalizability and reproducibility of TPS across centers and lesions are needed before TPS could be adopted as a national quality measure.

ICC = intraclass correlation coefficient   n = number of measurements
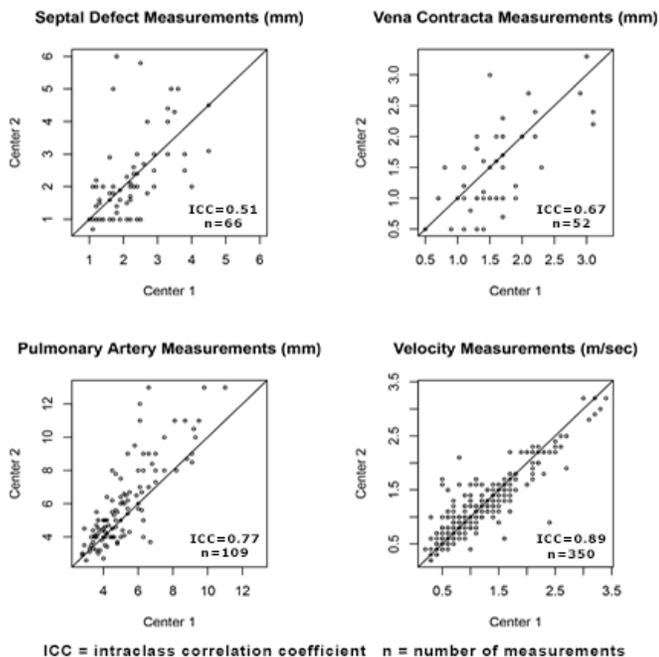
	Center 2		
	TPS Class 1	TPS Class 2	TPS Class 3
Center 1 TPS Class 1	26	7	0
TPS Class 2	1	16	2
TPS Class 3	0	4	11

Perfect agreement (along the diagonal from left to right) = 53
Disagreements Class 2 versus Class 3 = 6
TPS Class 1 (Optimal), TPS Class 2 (Adequate), TPS Class 3 (Inadequate)