Supplemental Online Content

Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol*. Published online February 18, 2021. doi:10.1001/jamacardio.2021.0185

This supplemental material has been provided by the authors to give readers additional information about their work.

eAppendix. Development of the AI-Guidance Algorithm

Background: The overall approach of this artificial intelligence (AI) echo guidance can be summarized as capturing the expertise of sonographers. Specifically, sonographers know what they are looking for, the diagnostic quality of what they are currently imaging, and know how to move the probe to acquire the desired imagery by examining the current ultrasound imagery. We used deep learning (DL), a form of AI, which is uniquely suited to capturing this perceptual ability.

DL is a form of machine learning based on artificial neural networks and refers to a number of techniques, but common attributes include composing simple computational elements into a layer, combining many layers into deep stacks, and adapting the parameters of the elements using supervised learning. While the theoretical underpinnings of deep learning have been around for decades, only recently have enough computational power, training data, and practical knowledge been available to achieve success.

DL has recently achieved a wide range of successes, including speech processing[1] and image understanding [2] as well as genomics[3] and medical imaging.[4] Our own work has indicated that deep learning algorithms for visual object recognition now rival the abilities of high-level primate cortex for rapid visual processing[5]. The ability of expert sonographers to interpret ultrasound in the "blink of an eye" is a similar form of visual discrimination, and capturing this ability with deep learning is the goal of this AI-guidance approach.

In order to help users to acquire high-quality echocardiograms, the guidance algorithm must be able to estimate the positioning of the ultrasound probe from the current imaging and provide real-time guidance based on that estimation for the user to arrive at the ideal positioning and image. It must first be able to grade the quality of the imagery and determine whether it meets diagnostic criteria. Therefore, the AI-guidance deep learning algorithm makes three simultaneous estimates: 1) diagnostic quality of the imagery, 2) six-dimensional (6D, probe position plus orientation) geometric distance between current probe location and a probe location anticipated to optimize the image, and 3) corrective probe manipulations to improve diagnostic quality. Importantly, the deep learning algorithm makes these estimates from only the ultrasound imagery; no trackers, fiducial markers, or additional sensors are needed for operation.

Training the network: As shown in **eFigure 1**, the deep learning algorithm was developed to guide operators to 10 diagnostic views of the heart: parasternal long axis view (PLAX); parasternal short axis view at aortic valve level (PSAX-AV), mitral valve level (PSAX-MV), and papillary muscle level (PSAX-PM); apical 2, 3, 4, 5 chamber views (Ap2, Ap3, Ap4, and Ap5); and subcostal 4 chamber (SC-4) and inferior vena cava view (SC-IVC). Fifteen (15) registered sonographers captured imagery on subjects over a range of BMI and clinical pathology to train the deep learning algorithm. This training dataset contained >5,000,000 individual data points on transducer location/orientation and observed imagery using ultrasound machines from multiple vendors. This dataset was curated and augmented with approximately 500,000 labels from expert sonographers and cardiologists annotating diagnostic quality and suggesting prescriptive guidance actions to improve the imaging (**eFigure 1 left**). Using this dataset, the algorithm was trained to estimate the 6D distance to each of the 10 views, define 81 unique prescriptive guidance actions, and diagnostic thresholds for each of the 10 views. The algorithm, consisting of more than 7,000,000 parameters, was trained using standard machine learning optimization techniques on a 31 tFLOP (trillion floating point operations per second) GPU array. Training the model took approximately 7.2 exaFLOP (1 exaFLOP is one million teraFLOPs or a quintillion (10^{18}) 32-bit multiplication floating point operations) over two weeks (**eFigure 1 center**). Further information on Caption Guidance may be found in the talk, "Development and Validation of a Breakthrough AI-Guided Echocardiography System" presented at the FDA-hosted public workshop titled, "Evolving Role of Artificial Intelligence in Radiological Imaging" held in Bethesda, Maryland on February 25 and 26, 2020, with slides [https://www.fda.gov/media/135734/download], and video [http://fda.yorkcast.com/webcast/Play/5ac1c24f9e48455c82011ab26837afad1d, with the relevant portion beginning at 42:10 of the video available.

Implementation: For real-time operation, the AI-guidance algorithm is implemented on a Terason uSmart 3200t Plus, which is a point-of-care ultrasound system with a phased-array transducer for cardiac imaging and an on-board P3000 NVIDIA GPU. The guidance implementation that runs on the Terason device as a relatively compact (1.5 GB) stand-alone executable and does not require any cloud connectivity or additional computational infrastructure. For real-time operation, the guidance algorithm is implemented in TensorFlow, an open source platform for machine learning originally developed by Google. The framerate of the ultrasound imagery is 30 fps with <20ms latency from ultrasound image formation to algorithm estimation, and the user is provided guidance updates 10 times per second. Note that even more compact implementations of the software have been developed for handheld ultrasound devices running on tablets and smartphones, which may enable use of this guidance on such devices in the future.

Detailed Description of Caption Guidance User Workflow: As shown in **eFigure 1 right**, the user interface is designed to aid medical professionals without prior ultrasound experience to perform diagnostic imaging. The user interface guides users through a predefined and customizable imaging workflow to capture a specific set of ultrasound views. In the clinical study, a protocol of 10 views was utilized. Each view was attempted sequentially according to the preset protocol. The user interface contains static guidance, which indicates the approximate starting location on the surface of the body, and a canonical image for the desired view in the protocol. Other relevant user interface features include the "quality meter," "prescriptive guidance," and "save best clip." Users were instructed to first observe the static guidance display to orient the probe to begin and to familiarize themselves with the desired view. They are then instructed to begin scanning and observe the response of the quality meter. Because the quality meter provides an estimate of the 6D distance from the desired view, the user can observe the response of the quality meter as it corresponds to their probe movements and choose to continue probe movements that increase the response of the quality meter. For example, moving the probe more medially may reduce the quality meter response, but moving more laterally may increase the response, thus giving the user feedback to continue with a lateral probe movement.

When the underlying algorithms detect a recognizable image appearance, users are presented with prescriptive guidance cues to guide probe movement with a specific motion. For example, an under-rotation may be detected for a parasternal long axis view acquisition, and the prescriptive guidance would then instruct the user to rotate slowly counterclockwise. When the user follows the recommended prescriptive guidance command, the quality meter typically responds with increasing response and users are instructed to continue their motion until they maximize the response of the quality meter.

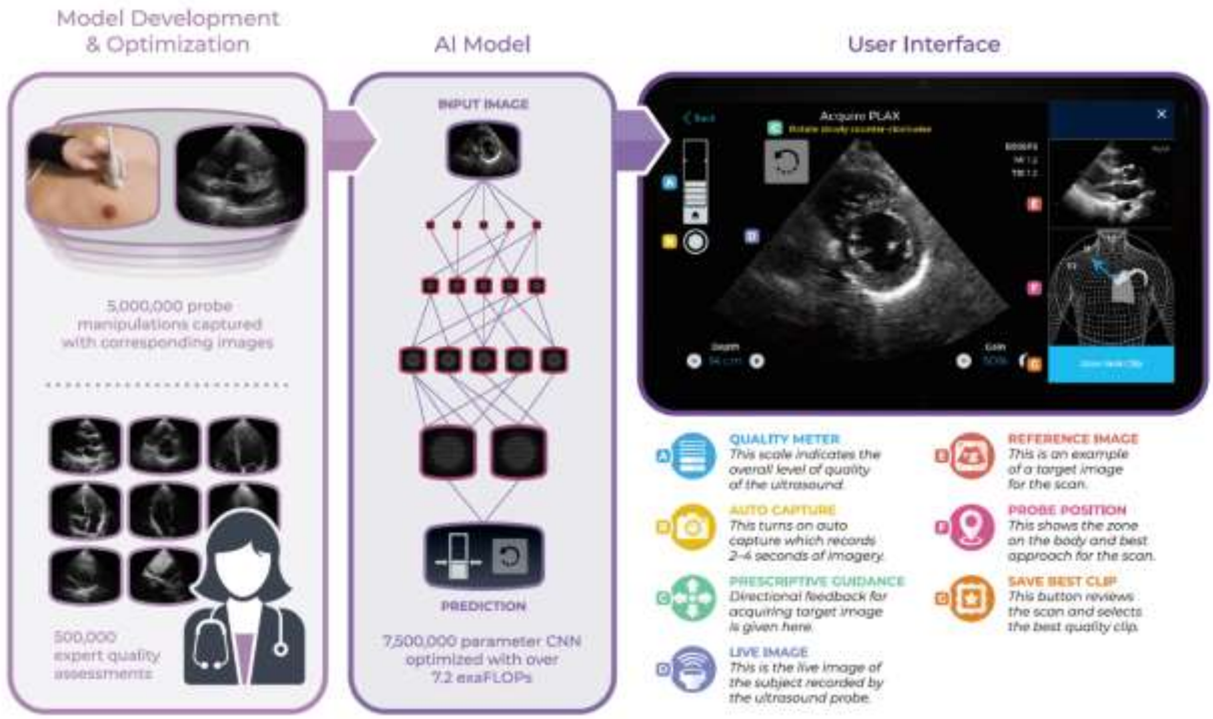
The quality meter also indicates a diagnostic quality threshold, and if the user maintains a quality meter level above this threshold, the software will automatically begin to capture a clip prospectively (called auto-capture) and store the clip as long as the user maintains the meter level above the threshold for at least 2 seconds up to 4 seconds; a clip shorter than 2 seconds will be discarded. This auto-captured 2 to 4 second clip is then utilized as the resulting clip for that view. If during scanning, the user does not cross the auto-record threshold within 2 minutes, the save best clip option will appear, enabling the user to proceed with the selection of the 2 second image sequence that produced the highest quality meter response over the 2 minutes. The user may either choose to tap on the save best clip option and proceed to the next view, or may continue to scan and attempt to achieve an auto-captured image. The save best clip feature was utilized in a high proportion of patient exams, most of which were deemed to be diagnostic quality even though they did not cross the auto-capture threshold. Note that the diagnostic threshold has been optimized to have high precision, rather than high recall, so as to optimize for high-diagnostic quality scans.

At completion of the 10 view protocol, users are presented a summary page that enables them to review the imagery they have acquired for each view. Upon review, users can elect to re-record a specific view, which then enters them into the workflow for that specific view before returning them to the summary page. Once review is complete, users chose to end the study and save the results. Studies were then transferred for storage and the clinical read using standard methods.

eFigure 2 shows in greater detail how an operator uses AI-guided feedback as they acquire each ultrasound view, including the process of following the Quality Meter and Prescriptive Guidance prompts to achieve Auto-Capture, and also a situation where the Save Best Clip is triggered. Through this process, operators obtain a video clip of the desired view by either achieving Auto-Capture or a Save Best Clip capture, repeating this process for each of the 10 views (which are shown in **eFigure 3)**.

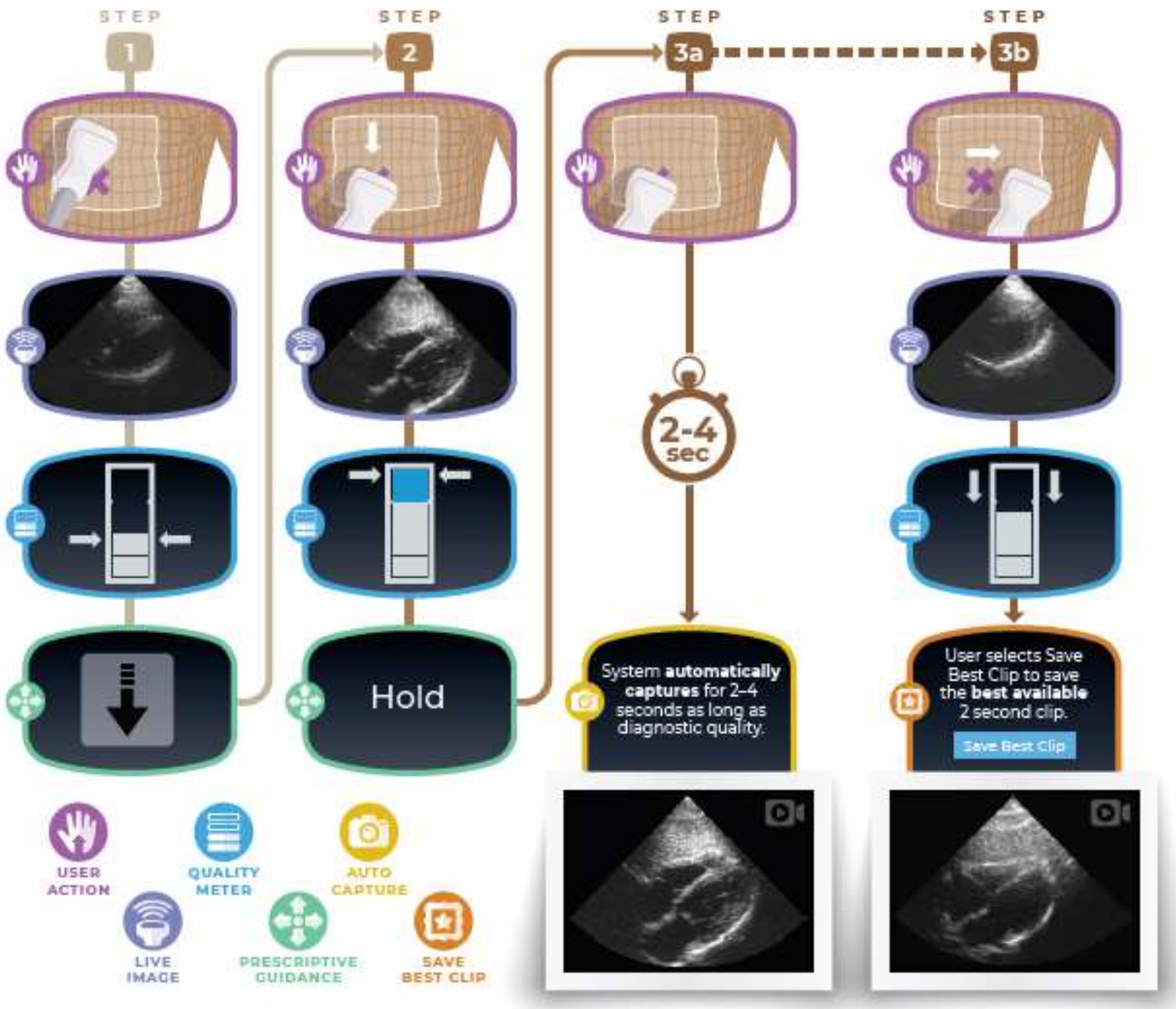
Testing / Validation: The individual components of the AI-guidance have been evaluated for the precision of estimates vs sonographer judgements, estimation of diagnostic quality, performance of prescriptive guidance cues, and pilot testing of novice user performance. In preparation for the prospective study described in this paper, we performed a pilot study of 4 nurses with no prior ultrasound experience. This study included 16 subjects with cardiac pathology and a range of body mass index (BMI). The pilot study produced numbers utilized in the power analyses we performed to determine the study size of this pivotal study. We obtained consistent results in the final study as compared to the pilot study. Additional descriptions of the component testing activities of neural network are beyond the scope of this paper.

In calculating the sample size for the current study, we recognized two sources of random variance in the data: the nurses and the patients. Accordingly, we approached this as a multi-reader, multi-case (MRMC) study (with the nurses serving as "readers" in this context): the MRMC approach does not assume that all nurses have the same skill level nor that all the patients present the same level of scanning "difficulty." Note that the variability of RN skill level can influence the precision of our estimate of acquisition success rate. Because the success/failure of the clinical trial depends not just on the level of acquisition success but the precision as well, we sized the study to sufficiently demonstrate the generalizability to our conclusions in a statistically significant manner. In particular, the study was powered to detect the primary endpoint's exceeding the performance goal of 80% (alpha = 0.05, beta = 0.2). This was done based on the result of the pilot study with the same study design (with a minor difference in initial RN training duration). The statistical power was estimated using iMRMC 4.0 software developed in the FDA, which provides the 95% CI around the point estimate for a given parameter and was used for the primary endpoint analysis. The software considers the variability of the performance of RNs as a source of variability as well as the variability of the outcomes across different patients as another source of variability (random effects model), performing multi-reader, multi-case analysis based on Gallas et al.[6] It was assumed that the mean effect size, the variance of RN performance, and the variance of success rates across patients would remain the same between the pilot study and the main study. As detailed in the manuscript, a study design with eight RNs performing 30 cases each led to a power of 0.92 for the sequential testing of the four primary endpoints with the above performance requirement.

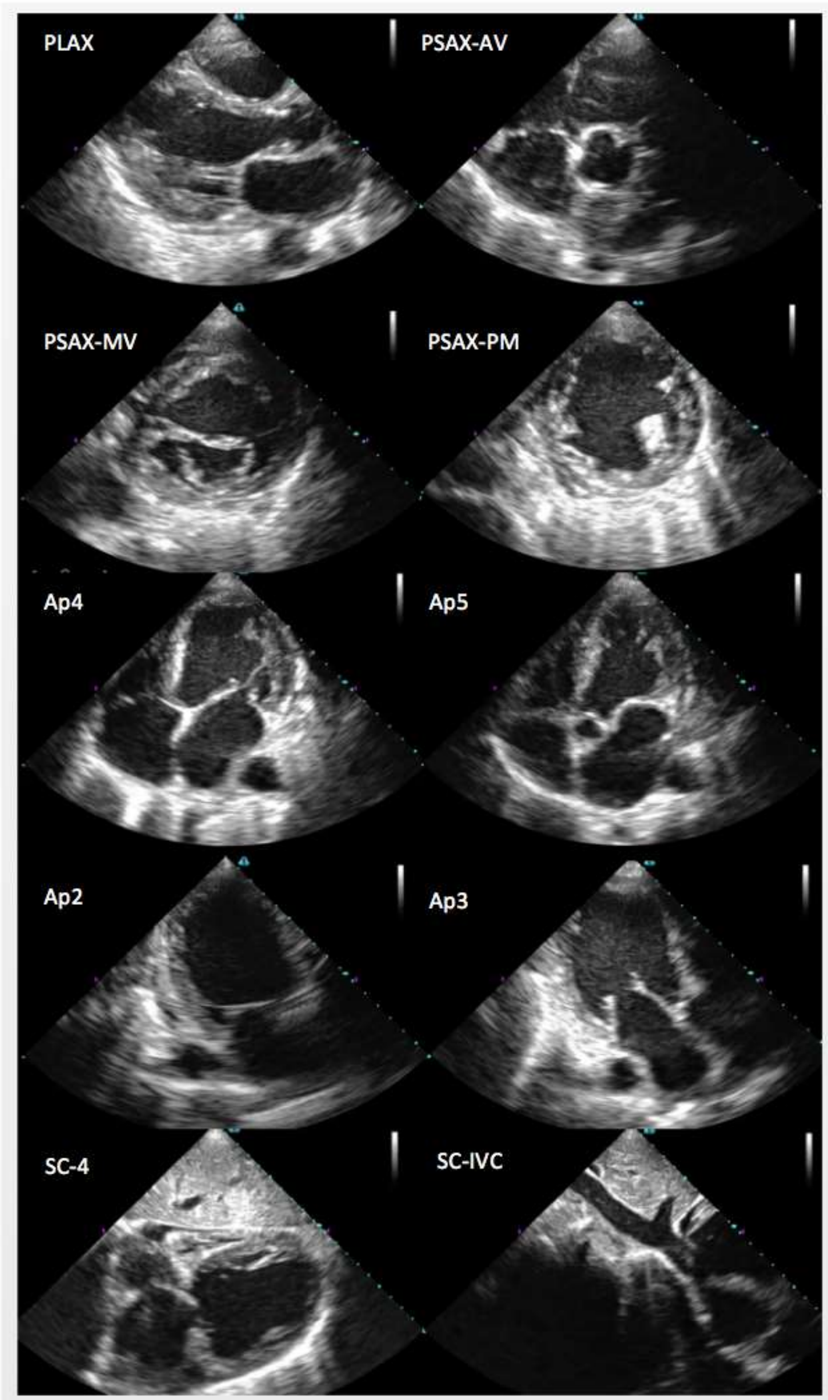


eFigure 1. Evidence base (left), neural network optimization (center), and user interface (right) of the AI guidance for echo acquisition.

Schematic diagram illustrating the deep learning algorithm training dataset, optimization, and runtime operation. To the **left** is shown how the impact on image appearance from millions of probe movements was captured along with hundreds of thousands of expert sonographer and cardiologist judgements of quality, as well as suggested manipulations to improve the image. This then was provided as the input training dataset to a multilayer convolutional neural network (**center**) to optimize the deep learning algorithm parameters using massive calculations on a 31 teraFLOPS (trillion floating point operations per second) GPU array running for two weeks (for a total of $7.2x10^{18}$ 32-bit calculations). To the **right** depicts the operation of the deep learning algorithm at runtime (during the operation by the nurses in the study). Note that during runtime the deep learning algorithm's only input is the live ultrasound image, and no positioning information or clinician input is necessary for the algorithm to judge quality and issue guidance commands. In the **right** panel, the guidance indicates that the user needs to "rotate [the probe] slowly counter-clockwise" in order to improve the parasternal long-axis image. Abbreviations: GPU, graphical processing unit; exaflops, 10^{18} floating point operations.



eFigure 2. The typical workflow for user interaction with the AI guidance.
Schematic diagram illustrating the user operation by the nurses during the study. Users begin (**Step 1**) by manipulating the probe position and watching for feedback from the quality meter. A guidance command may appear directing the user to make a specific probe manipulation to acquire a more diagnostic image. (**Step 2**) If the user follows the instruction, the guidance meter is likely to increase to a level appropriate for diagnostic purposes, as indicated by the quality meter. After holding the probe such that the quality meter remains in the diagnostic regime for sufficient time (**Step 3a**), the image will be auto recorded. If after a pre-specified time interval (**Step 3b**), this threshold has not been reached, the user may opt to capture the highest scoring clip thus far or continue scanning in hopes of achieving auto-capture.



eFigure 3. Ten representative still images acquired by a study nurse from a single patient.
Representative still images of 10 standard TTE views acquired by a nurse using the DL algorithm that were judged to be of diagnostic quality. Moving images are provided in the on-line supplement (**eVideo 2-11**). Abbreviations: PLAX, parasternal long-axis view; PSAX-AV, -MV, -PM, parasternal short axis view at the aortic valve, mitral valve, and papillary muscle

levels; Ap4, 5, 2, 3, apical 4, 5, 2, 3 chamber view, SC-4, subcostal 4-chamber view; SC-IVC, subcostal inferior vena cava view.

eTable 1. Demographics of Enrolled Patients

	Total
Total Enrolled	244
Sex, n (% of enrolled)	
Female	103 (42.2%)
Male	141 (57.8%)
Ethnicity, n (% of enrolled)	
Hispanic or Latino	7 (2.9%)
Not Hispanic or Latino	232 (95.1%)
Unknown/Not Reported	5 (2.0%)
Race, n (% of enrolled)	
White	187 (76.6%)
Black/ African American	43 (17.6%)
Asian	4 (1.6%)
American Indian / Alaska Native	2 (0.8%)
Unknown/ Not Reported	5 (2.0%)
Other	3 (1.2%)
Age (years), range	61.3 ± 15.67 (20-91)
Prior Cardiac Diagnoses	
Hypertension	149 (61.1%)
Hyperlipidemia	110(45.1%)
Diabetes	48 (19.7%)
Heart Failure	53 (21.%)
Atrial Fibrillation	62 (25.4%)
Other Arrythmias	36 (14.8%)
Coronary Artery Disease	75 (30.7%)
Prior Heart Attack	26 (10.7%)
Valvular Heart Disease	121 (49.6%)
Pulmonary Hypertension	12 (4.9%)
Heart Transplant	0
Cardiomyopathies	36 (14.3%)
Congenital Heart Disease	17 (7.0%)
Other	56 (23%)
None	23 (9.4%)
Not Reported	12 (4.9%)
Prior Non-Cardiac Diagnosis	
Renal Disease	32 (13.1%)
COPD/Emphysema	16 (6.6%)
Pulmonary Embolus	8 (3.3%)
Systemic infiltrative disease like amyloid or hemachromatosis	1 (0.4%)
Cancer	42 (17.2%)
Underwent Chemotherapy	15 (6.1%)
Underwent Radiation	11 (4.5%)
Other	44 (18%)
None	132 (54.1%)
Not Reported	2 (0.8%)

eTable 2. Summary of Patients with Cardiac Abnormalities Identified through Scheduled Standard-of-Care Echocardiogram by Study Site

Cardiac Abnormality Identified through Scheduled Echocardiogram	Study Site		All
	Northwestern (N=121) n (%)	Minneapolis (N=123) n (%)	
Patients with any cardiac abnormality	111 (91.7)	112 (91.1)	223 (91.4)
Abnormal left ventricular size or function	93 (76.9)	84 (68.3)	177 (72.5)
Abnormal right ventricular size or function	36 (29.8)	19 (15.4)	55 (22.5)
Abnormal left atrial size	50 (41.3)	52 (42.3)	102 (41.8)
Abnormal right atrial size	34 (28.1)	33 (26.8)	67 (27.5)
Septal defect	2 (1.7)	0 (0)	2 (0.8)
Abnormal mitral valve	86 (71.1)	65 (52.8)	151 (61.9)
Abnormal tricuspid valve	53 (43.8)	67 (54.5)	120 (49.2)
Abnormal aortic valve	54 (44.6)	58 (47.2)	112 (45.9)
Non-trivial pericardial effusion	6 (5.0)	2 (1.6)	8 (3.3)
Abnormal inferior vena cava size	0 (0)	6 (4.9)	6 (2.5)
Patent foramen ovale	1 (0.8)	0 (0)	1 (0.4)
Other abnormality	4 (3.3)	5 (4.1)	9 (3.7)
Implanted Medical Devices	22 (18.2)	32 (26.0)	54 (22.1)
Pacemaker/ICD	11 (9.1)	12 (9.8)	23 (9.4)
Leadless pacemaker	0	0	0
Prosthetic heart valve	10 (8.3)	20 (16.3)	30 (12.3)
LAA closure device	0	0	0
Atrial septal defect closure device	2 (1.7)	0	2 (0.8)
Ventricular septal defect closure device	0	1 (0.8)	1 (0.4)
Patent foramen ovale closure device	0	0	0
Valve repair device	6 (5.0)	4 (3.3)	10 (4.1)

eTable 3. Proportion of Nurse-Acquired EchoGPS Echocardiography of Sufficient Quality to Assess Clinical Parameters (Secondary Endpoints) in Nurse Scan Population.

Endpoint	#	Clinical Parameter	Performance Goal	Total Number Scans Performed	Number of Scans of Sufficient Quality	% of Scans of Sufficient Quality (95% CI)
Secondary Endpoints	**1**	Qualitative visual assessment of right ventricular function	N/A	240	219	91.3% (85.7%, 96.8%)
	2	Qualitative visual assessment of left atrial size	N/A	240	227	94.6% (90.7%, 98.5%)
	3	Qualitative visual assessment of aortic valve	N/A	240	220	91.7% (88.0%, 95.3%)
	4	Qualitative visual assessment of mitral valve	N/A	240	231	96.3% (93.9%, 98.6%)
	5	Qualitative visual assessment of tricuspid valve	N/A	240	200	83.3% (77.0%, 89.7%)
	6	Qualitative visual assessment of inferior vena cava size	N/A	240	138	57.5% (41.5%, 73.5%)

eTable 4. Performance of nurse scans for primary and secondary endpoints stratified by BMI and presence of cardiac pathology.

Endpoint	#	Clinical Parameter	BMI Category			Presence of Known Cardiac Abnormality at Time of Enrollment		Total (N=240) n (%) (95% CI)
			< 25 (N=85) n (%)	25 to < 30 (N=76) n (%)	≥ 30 (N=79) n (%)	Present (N=153) n (%)	Absent (N=87) n (%)	
Primary Endpoints	**1**	Qualitative Visual Assessment of Left Ventricular Size	84 (98.8%)	76 (100%)	77 (97.5%)	151 (98.7%)	86 (98.9%)	237 (98.8%) (96.7%, 100%)
	2	Qualitative visual assessment of global Left ventricular function	84 (98.8%)	76 (100%)	77 (97.5%)	151 (98.7%)	86 (98.9%)	237 (98.8%) (96.7%, 100%)
	3	Qualitative visual assessment of right ventricular size	85 (100%)	70 (92.1%)	67 (84.8%)	138 (90.2%)	84 (96.6%)	222 (92.5%) (88.1%, 96.9%)
	4	Qualitative visual assessment of non-trivial pericardial effusion	84 (98.8%)	76 (100%)	77 (97.5%)	152 (99.3%)	85 (97.7%)	237 (98.8%) (96.7%, 100%)
Secondary Endpoints	**5**	Qualitative visual assessment of right ventricular function	85 (100.0%)	70 (92.1%)	64 (81.0%)	135 (88.2%)	84 (96.6%)	219 (91.3%) (85.7%, 96.8%)
	6	Qualitative visual assessment of left atrial size	82 (96.5%)	73 (96.1%)	72 (91.1%)	143 (93.5%)	84 (96.6%)	227 (94.6%) (90.7%, 98.5%)
	7	Qualitative visual assessment of aortic valve	80 (94.1%)	73 (96.1%)	67 (84.8%)	137 (89.5%)	83 (95.4%)	220 (91.7%) (88.0%, 95.3%)
	8	Qualitative visual assessment of mitral valve	82 (96.5%)	74 (97.4%)	75 (94.9%)	146 (95.4%)	85 (97.7%)	231 (96.3%) (93.9%, 98.6%)
	9	Qualitative visual assessment of tricuspid valve	83 (97.6%)	62 (81.6%)	55 (69.6%)	123 (80.4%)	77 (88.5%)	200 (83.3%) (77.0%, 89.7%)
	10	Qualitative visual assessment of inferior vena cava size	62 (72.9%)	34 (44.7%)	42 (53.2%)	82 (53.6%)	56 (64.4%)	138 (57.5%) (41.5%, 73.5%)

eTable 5. Panel Variability: Extent of Agreement among Cardiologists in Rating Acceptability of Echocardiography for Clinical Parameter Assessment by Primary and Secondary Parameters

	All 5 out of 5 cardiologists agree		At least 4 out of 5 cardiologists agree		At least 3 out of 5 cardiologists agree	
	n	%	n	%	n	%
Nurse Exams (N=240)						
1 - Qualitative Visual Assessment of Left Ventricular Size	232	96.7%	238	99.2%	240	100.0%
2 - Qualitative Visual Assessment of Left Ventricular Global Function	231	96.3%	238	99.2%	240	100.0%
3 - Qualitative Visual Assessment of Right Ventricular Size	143	59.6%	214	89.2%	240	100.0%
4 - Qualitative Visual Assessment of Non-trivial Pericardial Effusion	220	91.7%	230	95.8%	240	100.0%
5 - Qualitative Visual Assessment of Inferior Vena Cava Size,	147	61.3%	192	80.0%	240	100.0%
6 - Qualitative Visual Assessment of Right Ventricular Function	141	58.8%	216	90.0%	240	100.0%
7 - Qualitative Visual Assessment of Left Atrial Size	206	85.8%	224	93.3%	240	100.0%
8 - Qualitative Visual Assessment of Aortic Valve	169	70.4%	216	90.0%	240	100.0%
9 - Qualitative Visual Assessment of Mitral Valve	224	93.3%	235	97.9%	240	100.0%
10 - Qualitative Visual Assessment of Tricuspid Valve	157	65.4%	209	87.1%	240	100.0%
Sonographer Exams (N=235)						
1 - Qualitative Visual Assessment of Left Ventricular Size	233	99.1%	235	100.0%	235	100.0%
2 - Qualitative Visual Assessment of Left Ventricular Global Function	232	98.7%	235	100.0%	235	100.0%
3 - Qualitative Visual Assessment of Right Ventricular Size	171	72.8%	218	92.8%	235	100.0%
4 - Qualitative Visual Assessment of Non-trivial Pericardial Effusion	229	97.4%	233	99.1%	235	100.0%
5 - Qualitative Visual Assessment of Inferior Vena Cava Size	167	71.1%	214	91.1%	235	100.0%
6 - Qualitative Visual Assessment of Right Ventricular Function	168	71.5%	220	93.6%	235	100.0%
7 - Qualitative Visual Assessment of Left Atrial Size	228	97.0%	232	98.7%	235	100.0%
8 - Qualitative Visual Assessment of Aortic Valve	219	93.2%	232	98.7%	235	100.0%
9 - Qualitative Visual Assessment of Mitral Valve	232	98.7%	234	99.6%	235	100.0%
10 - Qualitative Visual Assessment of Tricuspid Valve	175	74.5%	212	90.2%	235	100.0%

eTable 6. Acceptability of Nurse-Acquired Caption Guidance Echocardiography for Primary Clinical Parameter Assessment (Primary Endpoints) in Nurse Scan Population (N=240), by Sequence Number of Scan Within Nurse

#	Clinical Parameter	Sequence Number of Scans Within Nurse			All (N=240) n (%)
		1-10 (N=80) n (%)	11-20 (N=80) n (%)	21-30 (N=80) n (%)	
1	Qualitative Visual Assessment of Left Ventricular Size	79 (98.8%)	80 (100.0%)	78 (97.5%)	237 (98.8%)
2	Qualitative Visual Assessment of Global Left Ventricular Function	79 (98.8%)	80 (100.0%)	78 (97.5%)	237 (98.8%)
3	Qualitative Visual Assessment of Right Ventricular Size	76 (95.0%)	72 (90.0%)	74 (92.5%)	222 (92.5%)
4	Qualitative Visual Assessment of Non-Trivial Pericardial Effusion	79 (98.8%)	79 (98.8%)	79 (98.8%)	237 (98.8%)

eTable 7. Cross-Classification of Cardiologists' Clinical Assessment Using Nurse-Acquired vs. Sonographer-Acquired Echocardiograms - Primary Endpoints Qualitative Visual Assessment among Patients for Whom a Qualitative Visual Assessment Could Be Made in Both Scan Populations

	Sonographer-acquired echo w/o platform				
Clinical Parameter Assessed Nurse-acquired echo w/ Caption Guidance platform					**% Overall Agreement and C.I.**
1 - Qualitative Visual Assessment of Left Ventricular Size	Normal or Borderline	Abnormal (Enlarged)	No majority assessment among cardiologists	Total	
Normal or Borderline	203	8	1	212	
Abnormal (Enlarged)	1	19	0	20	
No majority assessment among cardiologists	0	0	0	0	
Total	204	27	1	232	95.7 (92.2, 97.6)
2 - Qualitative Visual Assessment of Left Ventricular Global Function	Normal or Borderline	Reduced (EF <=50)	No majority assessment among cardiologists	Total	
Normal or Borderline	187	5	0	192	
Reduced (EF <=50)	3	37	0	40	
No majority assessment among cardiologists	0	0	0	0	
Total	190	42	0	232	96.6 (93.3, 98.2)
3 - Qualitative Visual Assessment of Right Ventricular Size	Normal or Borderline	Abnormal (Enlarged)	No majority assessment among cardiologists	Total	
Normal or Borderline	185	1	3	189	
Abnormal (Enlarged)	5	11	2	18	
No majority assessment among cardiologists	5	0	1	6	
Total	195	12	6	213	92.5 (88.1, 95.3)
4 - Qualitative Visual Assessment of Non-trivial Pericardial Effusion	Absent	Present	No majority assessment among cardiologists	Total	
Absent	227	0	0	227	
Present	1	3	0	4	
No majority assessment among cardiologists	0	0	0	0	
Total	228	3	0	231	99.6 (97.6, 99.9)

5 - Qualitative Visual Assessment of Right Ventricular Function	Normal or Borderline	Reduced	No majority assessment among cardiologists	Total	
Normal or Borderline	184	0	3	187	
Reduced	4	12	2	18	
No majority assessment among cardiologists	5	1	1	7	
Total	193	13	6	212	92.9 (88.7, 95.7)

6 - Qualitative Visual Assessment of Left Atrial Size	Normal	Enlarged	No majority assessment among cardiologists	Total	
Normal	160	18	1	179	
Enlarged	8	32	0	40	
No majority assessment among cardiologists	0	2	1	3	
Total	168	52	2	222	86.9 (81.9, 90.7)

7 - Qualitative Visual Assessment of Aortic Valve	Structurally normal	Structurally abnormal	Suspected device	No majority assessment among cardiologists	Total	
Structurally normal	172	2	1	0	175	
Structurally abnormal	4	11	0	0	15	
Suspected device	1	0	9	2	12	
No majority assessment among cardiologists	4	4	2	1	11	
Total	181	17	12	3	213	90.6 (85.9, 93.8)

8 - Qualitative Visual Assessment of Mitral Valve	Structurally normal	Structurally abnormal	Suspected device	No majority assessment among cardiologists	Total	
Structurally normal	198	3	1	1	203	
Structurally abnormal	4	8	2	2	16	
Suspected device	0	0	4	0	4	
No majority assessment among cardiologists	1	1	0	0	2	
Total	203	12	7	3	225	93.3 (89.3, 95.9)

9 - Qualitative Visual Assessment of Tricuspid Valve	Structurally normal	Structurally abnormal	Suspected device	No majority assessment among cardiologists	Total	
Structurally normal	171	0	4	0	175	
Structurally abnormal	1	0	0	0	1	
Suspected device	1	0	5	0	6	
No majority assessment among cardiologists	1	0	2	1	4	
Total	174	0	11	1	186	95.2 (91.1, 97.4)

10 - Qualitative Visual Assessment of Inferior Vena Cava Size	Normal	Dilated	No majority assessment among cardiologists	Total	
Normal	103	6	3	112	
Dilated	5	6	3	14	
No majority assessment among cardiologists	4	1	0	5	
Total	112	13	6	131	83.2 (75.9, 88.6)

eReferences.

1. Hinton G, Deng L, Yu D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition. In. Signal Processing Magazine2012.
2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural net-works2012, NIPS.
3. Xiong HY, Alipanahi B, Lee LJ, et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science. 2015;347(6218):1254806.
4. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv. 2013;16(Pt 2):411-418.
5. Cadieu CF, Hong H, Yamins DL, et al. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput Biol. 2014;10(12):e1003963.
6. Gallas BD, Pennello GA, Myers KJ. Multireader multicase variance analysis for binary data. J Opt Soc Am A Opt Image Sci Vis. 2007;24(12):B70-80.