EchoNet-RCT: Blinded, Randomized Controlled Trial of Sonographer vs. Artificial Intelligence Assessment of Cardiac Function

Safety and Efficacy Study of AI LVEF

David Ouyang, MD

on behalf of EchoNet-RCT investigators

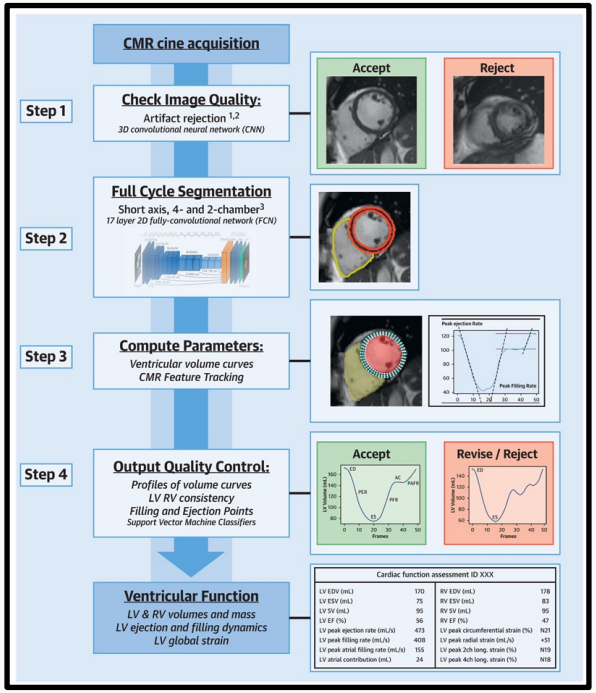
August 27, 2022

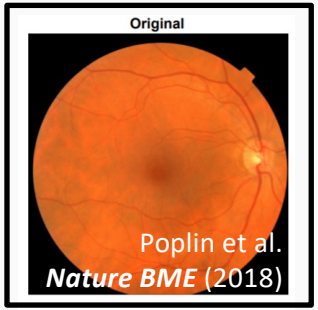
Disclosures

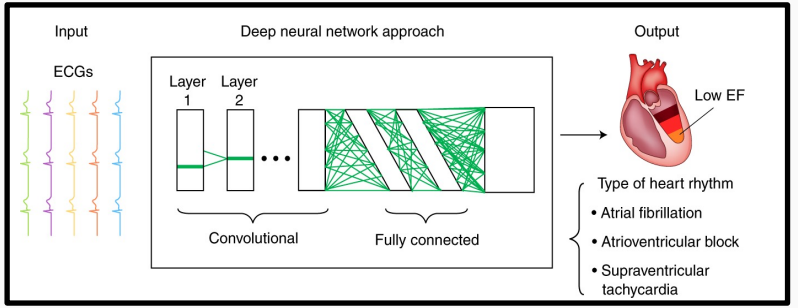
- Investigator initiated trial
- No industry funding for trial
- BH, JYZ, DO has patent for methods disclosed in Nature 2020

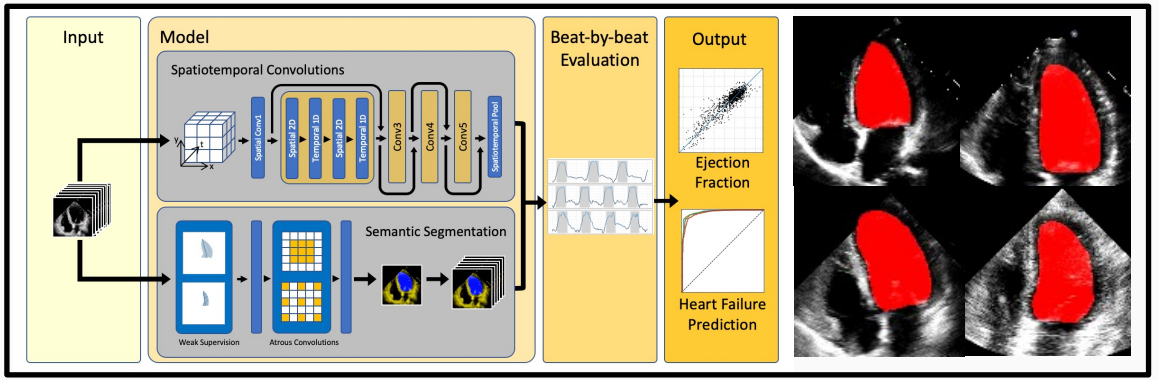
There has been tremendous progress in applying AI to cardiology, but no blinded randomized studies.

Attia et al. ***Nature Medicine*** (2019)





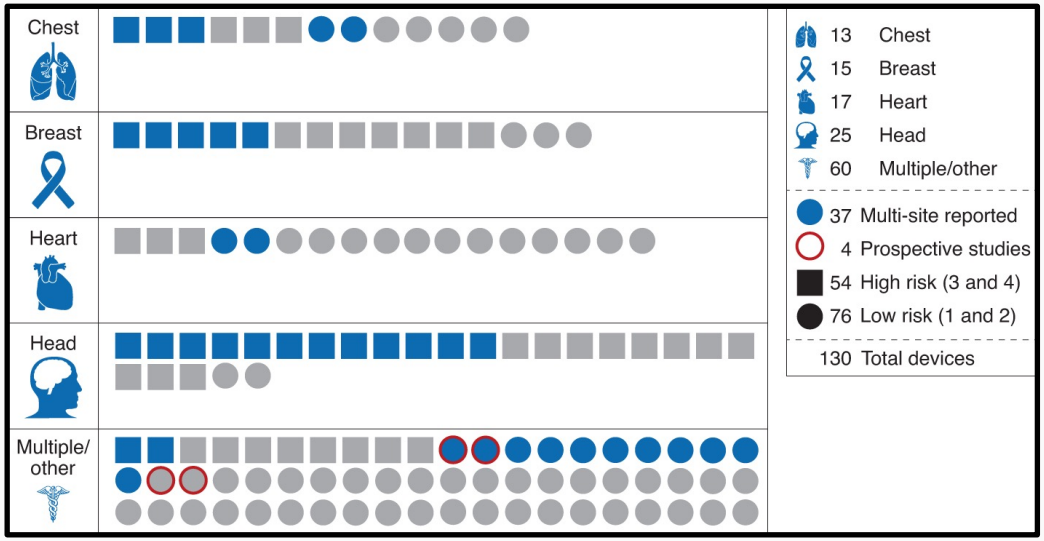




Ruijsink et al. ***JACC CV Img*** (2019)

Ouyang et al. ***Nature*** (2020)

For AI technologies, FDA 510k clearance and CE Mark does not currently require prospective clinical trials.



Wu et al. ***Nature Medicine*** (2021)

98% only retrospective data

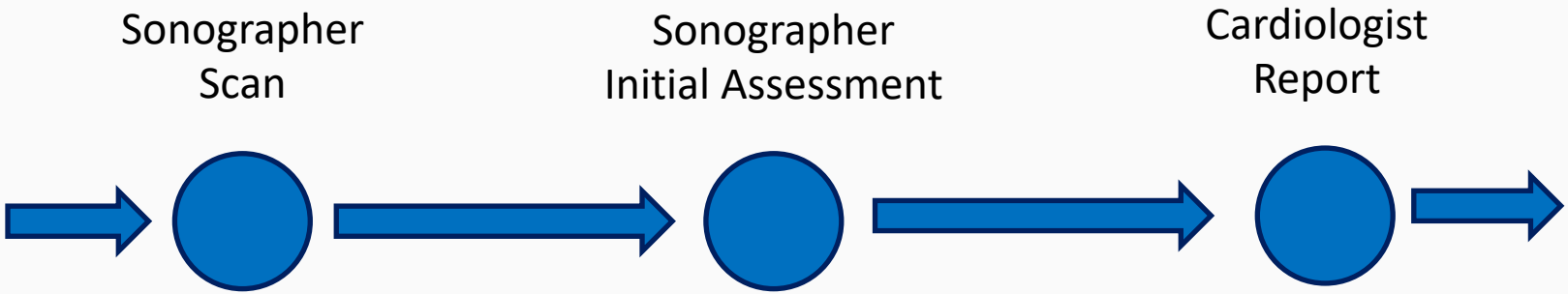
75% single site studies

None per randomized

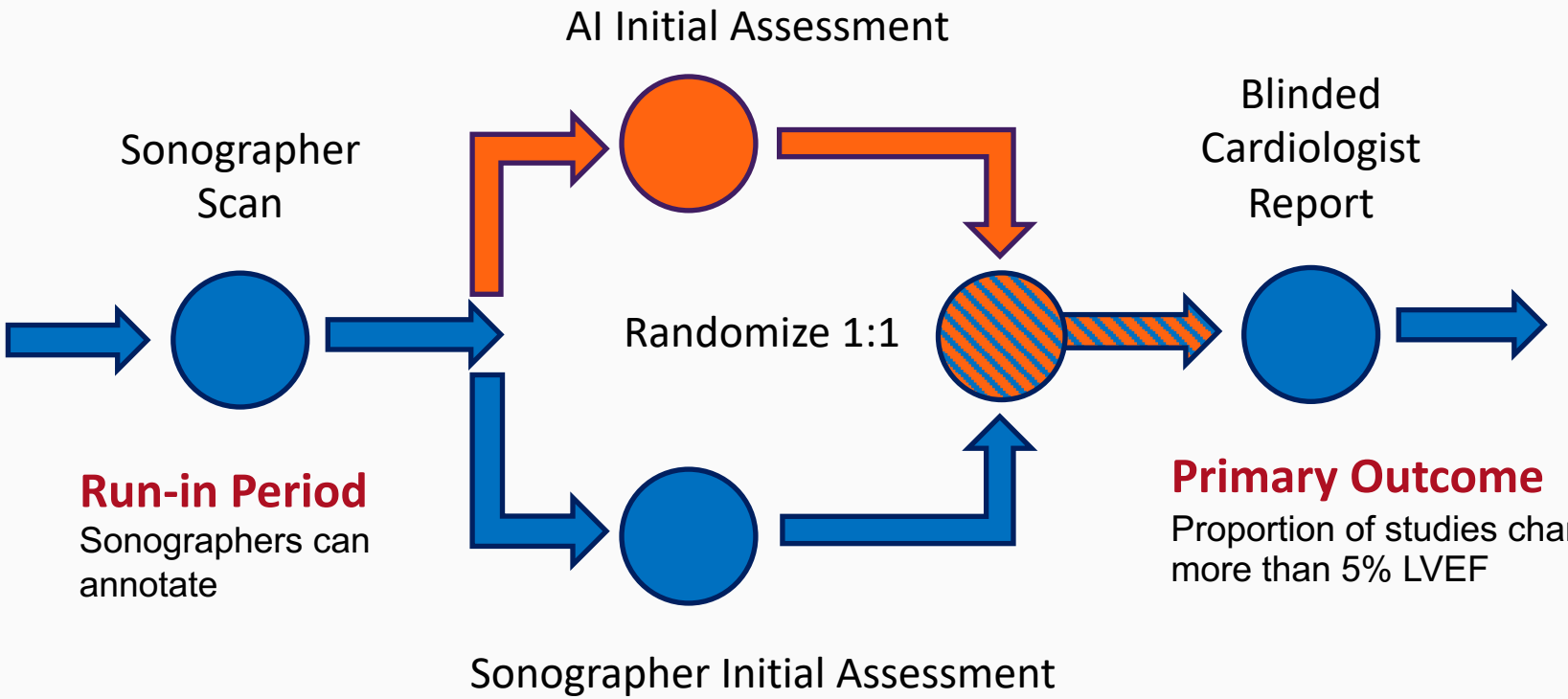
None are blinded

Blinding and randomization are core principles in clinical trials

Standard Clinical Workflow



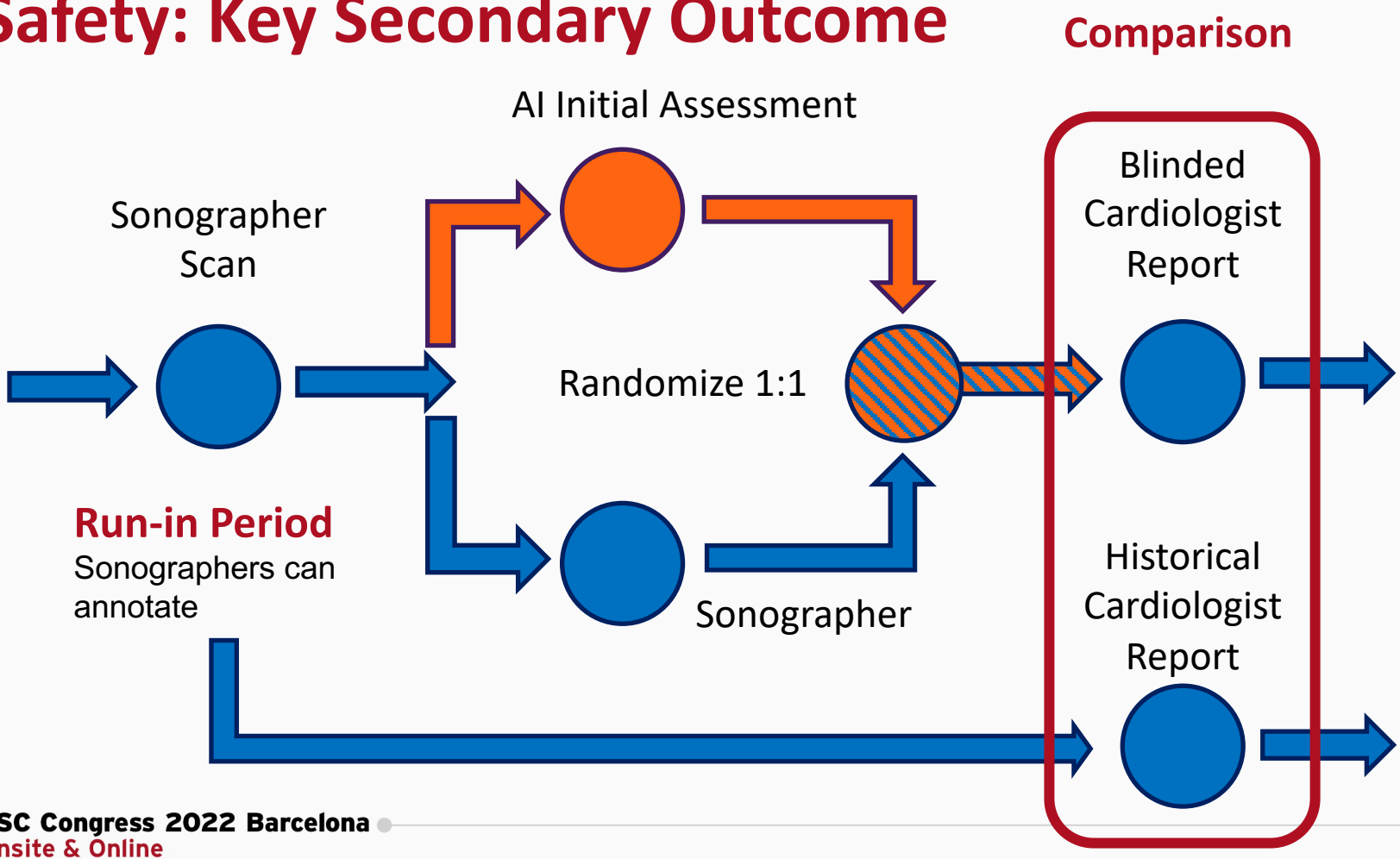
Trial Design



Trial Design

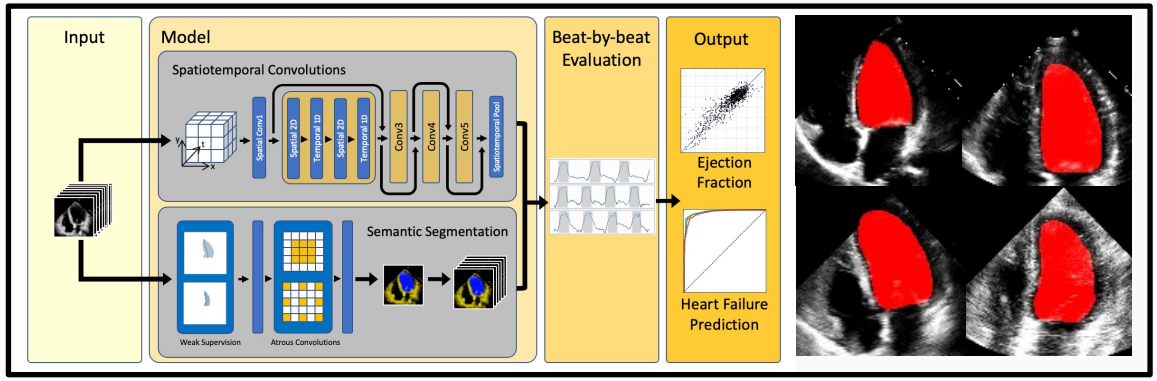
- Inclusion criteria: Adult transthoracic echocardiogram
- Exclusion criteria: Complete run-in period that sonographers can annotate
- **Primary Outcome:** Frequency and degree of change from initial (AI vs. sonographer) assessment to final cardiologist assessment
  - Substantial change defined as more than 5% LVEF
  - Non-inferiority Design: 8% vs. 5%, alpha of 0.05 and power of 0.9
    - 2834 studies needed, pre-planned to enroll 3500 studies as buffer against dropout
- **Secondary Outcomes:**
  - Sonographer Time
  - Cardiologist Time
  - Cardiologist Prediction of Agent of Initial Assessment
  - Change from Historical Cardiologist Assessment

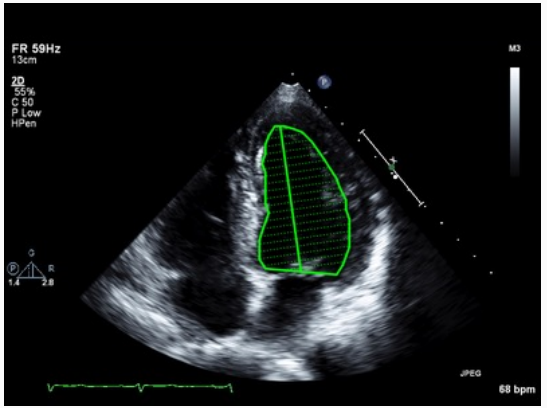
Safety: Key Secondary Outcome

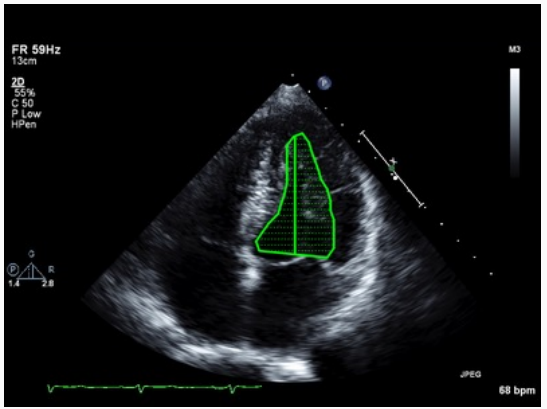


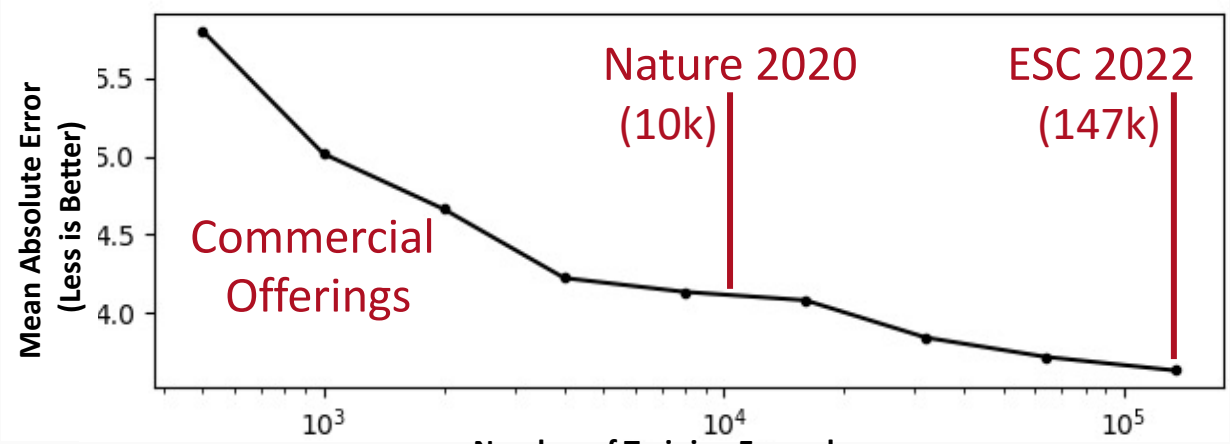
AI Model Design

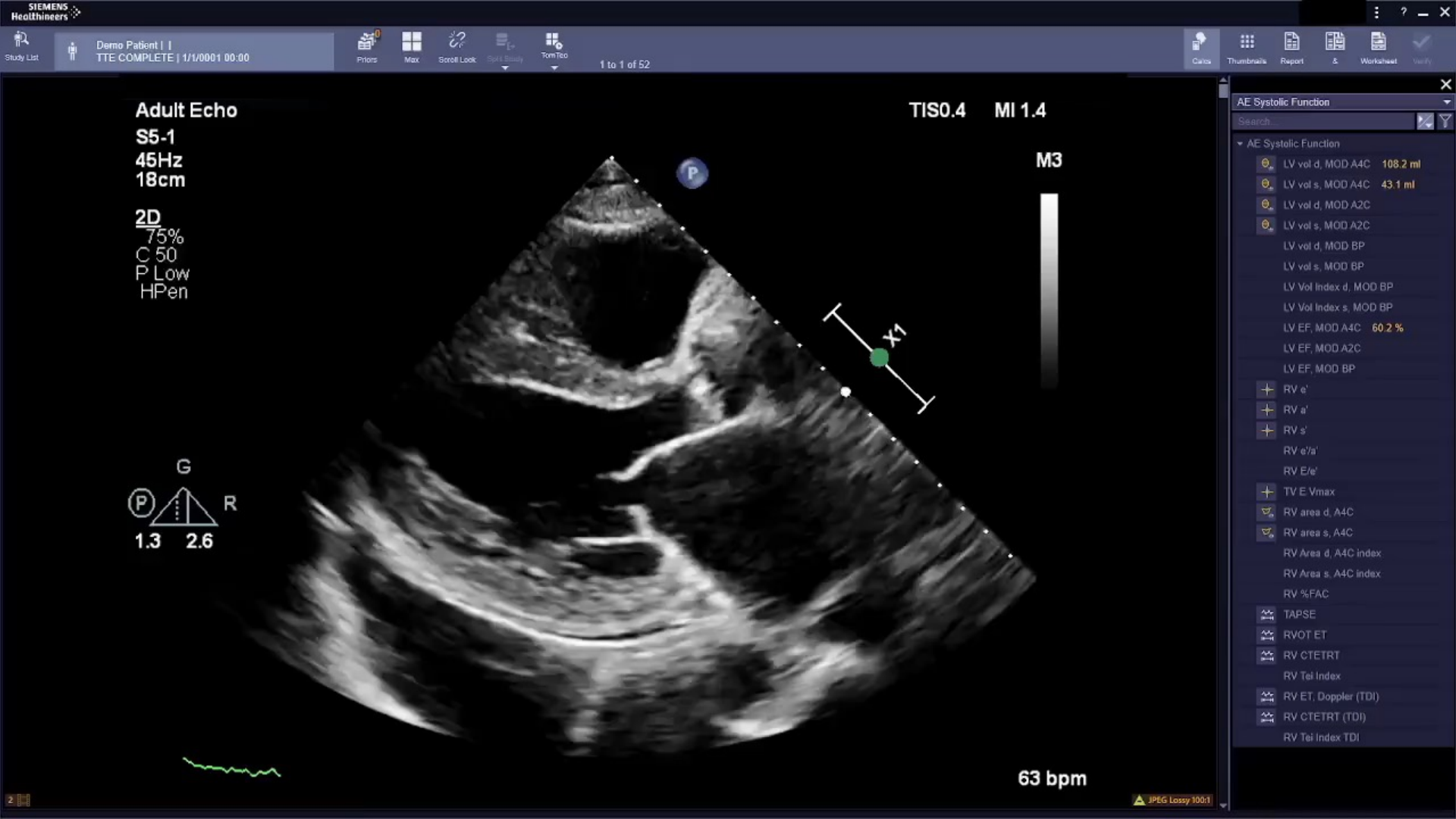
Ouyang et al. ***Nature*** (2020)










SIEMENS Healthineers
Study List
Demo Patient | |
TTE COMPLETE | 1/1/0001 00:00
Priors
Max
Scroll Look
TomTec
1 to 1 of 52
Calcs
Thumbnails
Report
Worksheet
Adult Echo
S5-1
45Hz
18cm
2D
75%
C 50
P Low
HPen
TIS0.4
MI 1.4
M3
X1
G
P
R
1.3
2.6
63 bpm
JPEG Lossy 100:1
AE Systolic Function
Search...
AE Systolic Function
LV vol d, MOD A4C 108.2 ml
LV vol s, MOD A4C 43.1 ml
LV vol d, MOD A2C
LV vol s, MOD A2C
LV vol d, MOD BP
LV vol s, MOD BP
LV Vol Index d, MOD BP
LV Vol Index s, MOD BP
LV EF, MOD A4C 60.2 %
LV EF, MOD A2C
LV EF, MOD BP
RV e'
RV a'
RV s'
RV e'/a'
RV E/e'
TV E Vmax
RV area d, A4C
RV area s, A4C
RV Area d, A4C index
RV Area s, A4C index
RV %FAC
TAPSE
RVOT ET
RV CTETRT
RV Tei Index
RV ET, Doppler (TDI)
RV CTETRT (TDI)
RV Tei Index TDI

Baseline Characteristics and Trial Flow

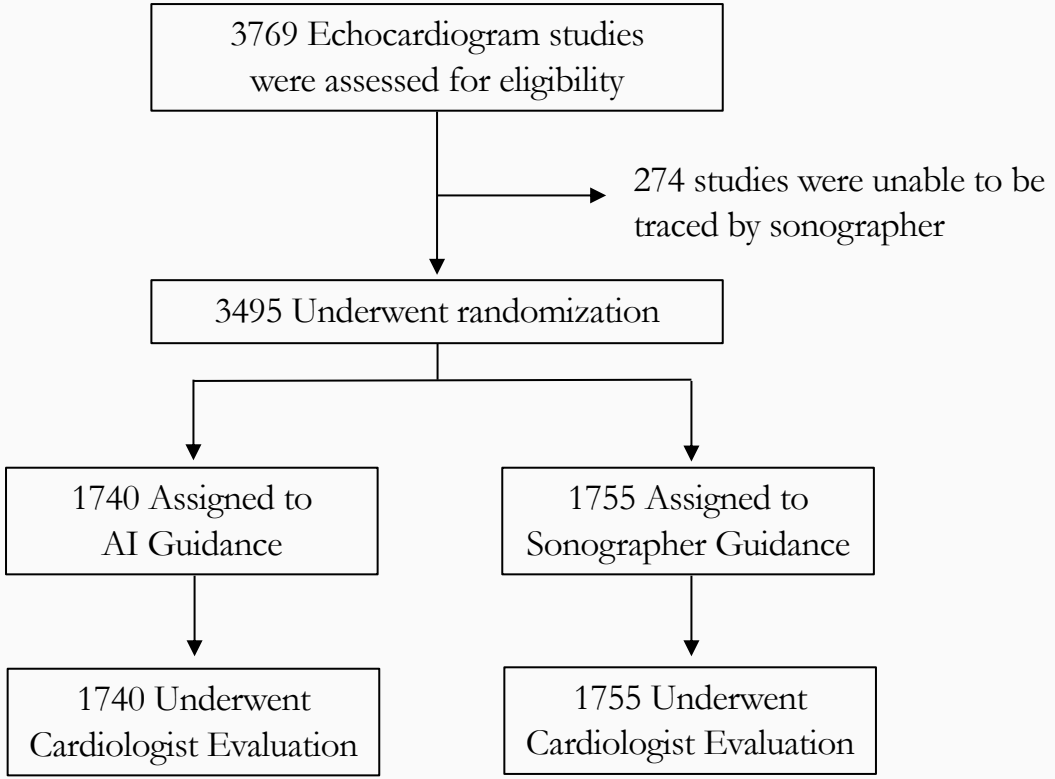
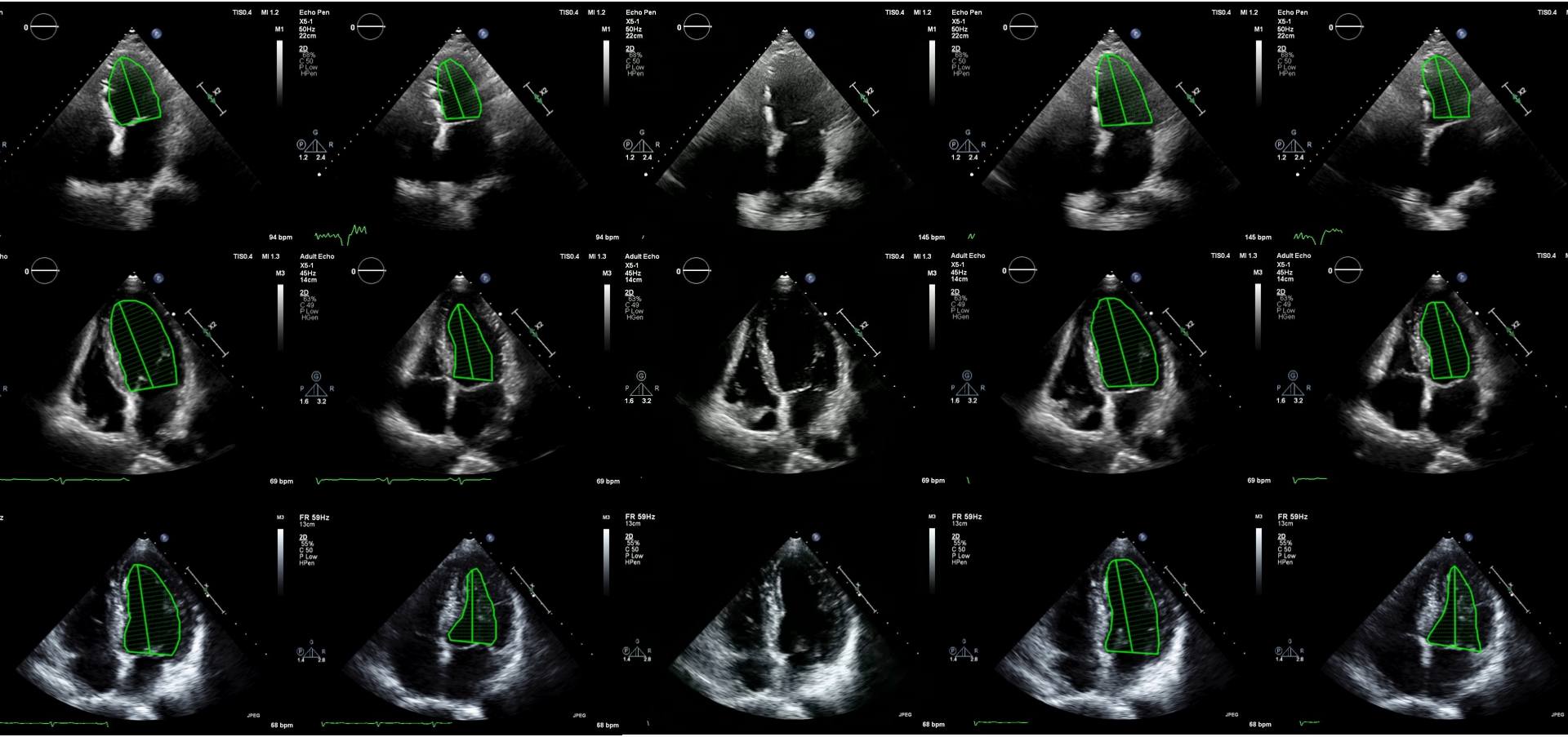
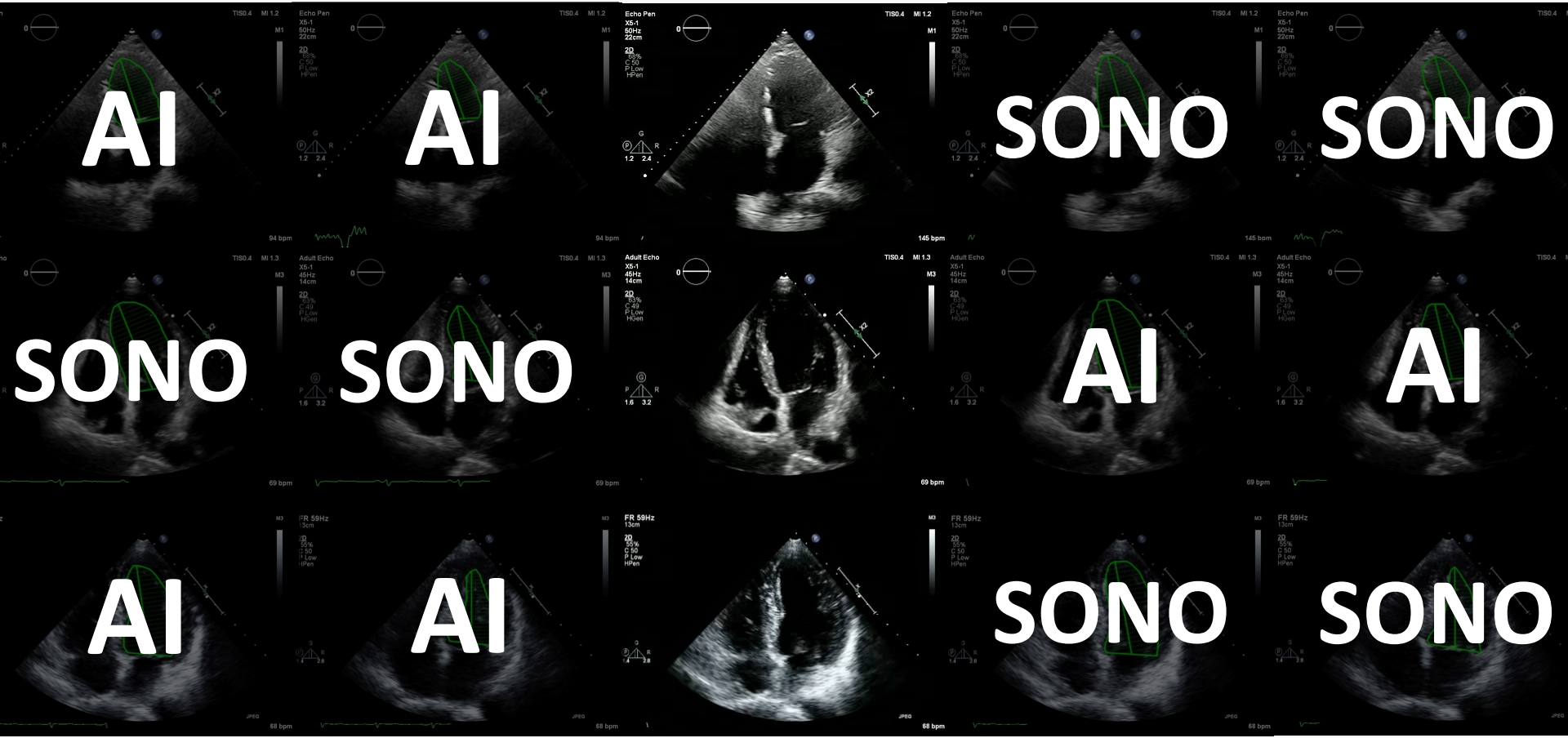


Table 1: Demographic and Imaging Study Characteristics.

Characteristic	Total (n = 3495)	AI (n = 1740)	Sonographer (n = 1755)
Age - yr	66.3±17.0	66.1±16.8	66.6±17.1
Sex - no. (%)			
Male	1983 (57%)	982 (56%)	1001 (57%)
Female	1512 (43%)	758 (44%)	754 (43%)
Race - no. (%)			
Non-Hispanic White	2041 (58%)	1032 (59%)	1009 (57%)
Black	479 (14%)	230 (13%)	249 (14%)
Hispanic	405 (12%)	203 (12%)	202 (12%)
Asian	273 (8%)	123 (7%)	150 (9%)
Other	237 (7%)	120 (7%)	117 (7%)
Unknown	38 (1%)	20 (1%)	18 (1%)
Pacific Islander	14 (0%)	8 (0%)	6 (0%)
American Indian	8 (0%)	4 (0%)	4 (0%)
Body Mass Index*	26.5±6.3	26.6±6.3	26.5±6.2
Comorbidities - no. (%)			
Hypertension	2019 (58%)	990 (57%)	1029 (59%)
Diabetes	884 (25%)	441 (25%)	443 (25%)
Coronary Artery Disease	1099 (31%)	547 (31%)	552 (31%)
Chronic Kidney Disease	882 (25%)	460 (26%)	422 (24%)
Atrial Fibrillation	867 (25%)	450 (26%)	417 (24%)
Prior Stroke	459 (13%)	225 (13%)	234 (13%)
Prior Clinical EF	58.1±14.3	58.1±14.2	58.0±14.4
Method of LVEF Evaluation - no. (%)			
Single Plane (A4C)	2249 (64%)	1107 (64%)	1142 (65%)
Biplane	1246 (36%)	633 (36%)	613 (35%)
Study Quality - no. (%)			
Poor	648 (19%)	314 (18%)	334 (19%)
Adequate	1725 (49%)	875 (50%)	850 (48%)
Good	236 (7%)	114 (7%)	122 (7%)
Not Specified	886 (25%)	437 (25%)	449 (26%)
Location - no. (%)			
Inpatient	2067 (59%)	1033 (59%)	1034 (59%)
Outpatient	1428 (41%)	707 (41%)	721 (41%)

A4C = Apical-4-Chamber, *BMI missing in 52 studies


TIS0.4 MI 1.2
Echo Pen
X5-1
50Hz
22cm
2D
95%
C 50
P Low
HPen
94 bpm
145 bpm
TIS0.4 MI 1.3
Adult Echo
X5-1
45Hz
14cm
2D
83%
C 49
P Low
HGen
69 bpm
FR 59Hz
13cm
2D
95%
C 50
P Low
HPen
68 bpm
JPEG


AI
AI
SONO
SONO
SONO
SONO
AI
AI
AI
AI
SONO
SONO

Cardiologists could not distinguish between AI and sonographer initial assessments

Correct	1130	(32.3%)
Unsure	1520	(43.4%)
Incorrect	845	(24.2%)

Bang's Blinding Index: 0.088
Between -0.2 to 0.2 is considered good blinding.

Trial Results

Outcome	AI (n = 1740)	Sonographer (n = 1755)	Mean Difference (95% CI)	P value
Primary Efficacy Outcome: Initial vs. Final Assessment				
Substantial Change	292 (16.8%)	478 (27.2%)	-10.5% (-13.2% to -7.7%)	< 0.001*
Mean Absolute Difference	2.79±5.53	3.77±5.22	-0.97 (-1.31 to -0.61)	< 0.001

* For both non-inferiority and superiority tests, all other tests were for superiority

Trial Results

Outcome	AI (n = 1740)	Sonographer (n = 1755)	Mean Difference (95% CI)	P value
Primary Efficacy Outcome: Initial vs. Final Assessment				
Substantial Change	292 (16.8%)	478 (27.2%)	-10.5% (-13.2% to -7.7%)	< 0.001*
Mean Absolute Difference	2.79±5.53	3.77±5.22	-0.97 (-1.31 to -0.61)	< 0.001
Key Secondary Safety Outcome: Final vs. Historical Cardiologist Assessment				
Substantial Change	871 (50.1%)	957 (54.5%)	-4.5% (-7.8% to -1.2%)	0.008
Mean Absolute Difference	6.29±5.94	7.23±6.18	-0.94 (-1.34 to -0.54)	< 0.001

* For both non-inferiority and superiority tests, all other tests were for superiority

Trial Results

Outcome	AI (n = 1740)	Sonographer (n = 1755)	Mean Difference (95% CI)	P value
Primary Efficacy Outcome: Initial vs. Final Assessment				
Substantial Change	292 (16.8%)	478 (27.2%)	-10.5% (-13.2% to -7.7%)	< 0.001*
Mean Absolute Difference	2.79±5.53	3.77±5.22	-0.97 (-1.31 to -0.61)	< 0.001
Key Secondary Safety Outcome: Final vs. Historical Cardiologist Assessment				
Substantial Change	871 (50.1%)	957 (54.5%)	-4.5% (-7.8% to -1.2%)	0.008
Mean Absolute Difference	6.29±5.94	7.23±6.18	-0.94 (-1.34 to -0.54)	< 0.001
Other Secondary Outcomes				
Sonographer time (s), median (IQR)	0 (0 - 0)	119 (77 - 173)	-131 (-134 to -127)	< 0.001
Cardiologist time (s), median (IQR)	54 (31 - 95)	64 (36 - 108)	-8 (-12 to -4)	< 0.001
Any Change	1100 (63.2%)	1218 (69.4%)	-6.2% (-9.3% to -3.1%)	< 0.001

* For both non-inferiority and superiority tests, all other tests were for superiority

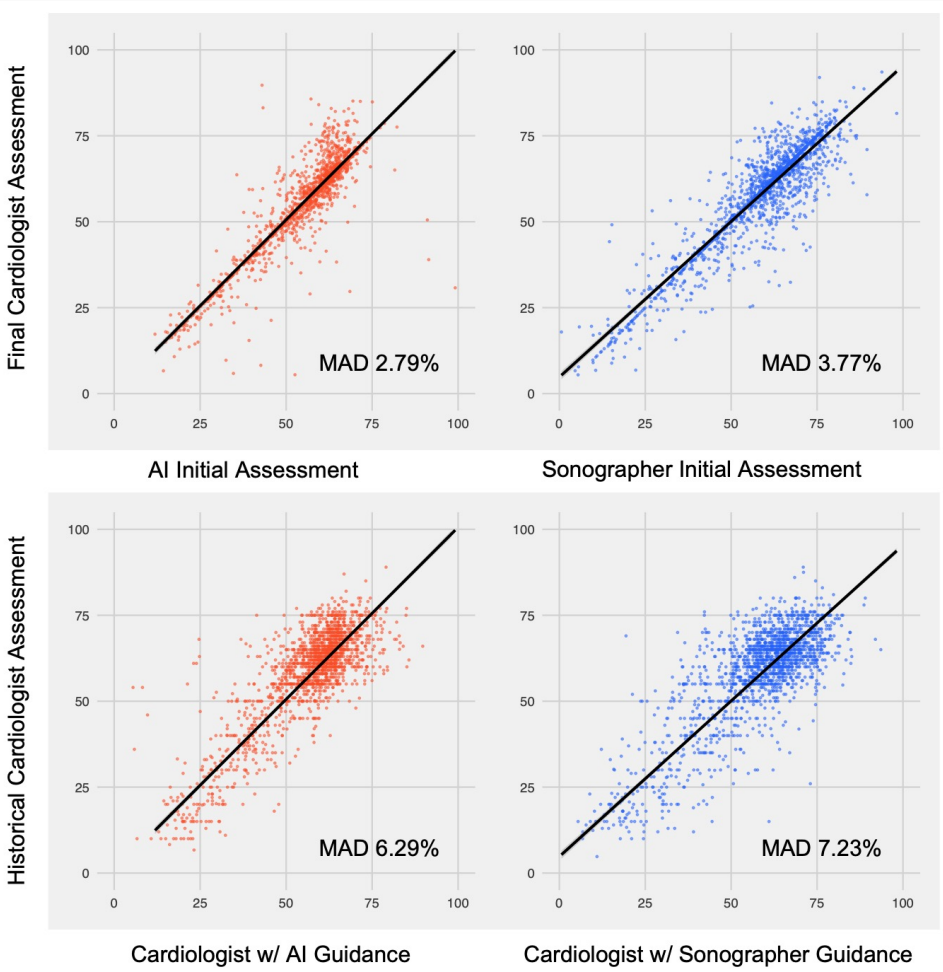
Trial Results

Primary Outcome

degree of change from initial (AI vs. sonographer) assessment to final cardiologist assessment

Key Secondary Outcome

degree of change from final cardiologist assessment to historical cardiologist assessment



Subgroup Analysis

Consistent results based on subgroups of patient characteristics, imaging study characteristics, and cardiologist prediction.

Subgroup	AI n	AI MAD	Sonographer n	Sonographer MAD	Difference (95% CI)
Method of LVEF Evaluation					
Single Plane	1107	3.20±6.41	1142	4.38±5.75	-1.19 (-1.69 to -0.68)
Biplane	633	2.09±3.37	613	2.61±3.77	-0.52 (-0.92 to -0.11)
Race					
White	1032	2.58±5.36	1009	3.71±5.18	-1.13 (-1.57 to -0.66)
Black	230	3.91±7.59	249	4.00±5.27	-0.08 (-1.22 to 1.14)
Hispanic	203	2.44±4.26	202	3.39±4.57	-0.95 (-1.82 to -0.09)
Asian	123	3.11±5.44	150	4.29±6.04	-1.18 (-2.55 to -0.23)
Other	152	2.77±4.11	145	3.74±5.26	-0.97 (-2.07 to -0.08)
Sex					
Male	982	2.75±5.92	1001	3.67±5.18	-0.92 (-1.40 to -0.42)
Female	758	2.85±4.97	754	3.89±5.26	-1.04 (-1.56 to -0.52)
Image Quality					
Poor	314	4.22±7.12	334	4.27±5.92	-0.05 (-1.04 to 0.97)
Adequate	875	2.45±5.36	850	3.53±5.01	-1.08 (-1.56 to -0.58)
Good	114	2.01±3.15	122	3.51±5.11	-1.51 (-2.62 to -0.45)
Not Specified	437	2.66±4.82	449	3.90±5.02	-1.24 (-1.89 to -0.58)
Location					
Inpatient	1033	3.09±5.59	1034	4.01±5.49	-0.92 (-1.40 to -0.45)
Outpatient	707	2.36±5.41	721	3.42±4.78	-1.05 (-1.57 to -0.51)
Cardiologist Prediction of Group					
AI	557	3.64±6.42	418	3.82±5.09	-0.18 (-0.91 to -0.54)
Sonographer	427	3.38±4.95	573	4.00±4.62	-0.62 (-1.21 to 0.00)
Uncertain	756	1.85±4.95	764	3.56±5.68	-1.72 (-2.26 to -1.17)

MAD = Mean Absolute Difference, LVEF = Left Ventricular Ejection Fraction

Limitations

- Single center
- In order to blind, AI was weakened
- No comparison with cross-sectional imaging
- Further work needed to assess long term impact

Strengths

- External validation of an AI model with publicly available code and representative training data
- Randomization with active comparator
- Blinding was quite successful
- Largest study of clinician test-retest of LVEF

Conclusion

- For adult patients undergoing echocardiographic quantification of cardiac function, initial assessment of LVEF by AI was noninferior and superior to initial sonographer assessment.

- After blinded review of initial LVEF assessment, cardiologists were less likely to substantially change their final report with initial AI assessment than sonographer assessment.

- AI guided assessment took less time for cardiologists to overread and was more consistent with historical cardiologist assessment (test-retest precision).

Thank you!

EchoNet-RCT Investigators

- Bryan He
- Alan C. Kwan, MD
- Jae Hyung Cho, MD, PhD
- Neal Yuan, MD
- Charles Pollick, MD
- Takahiro Shiota, MD
- Joseph Ebinger, MD
- Natalie A. Bello, MD
- Janet Wei, MD
- Kiranbir Josan, MD
- Grant Duffy
- Melvin Jujjavarapu
- Robert Siegel, MD
- Susan Cheng, MD
- James Y. Zou, PhD
- David Ouyang, MD