EchoNet-Core Lab Study: Blinded, Randomized Controlled Trial of Artificial Intelligence Guided Precision Assessment of Cardiac Function

~~August 31~~September 22, 2021

David Ouyang, MD

Smidt Heart Institute

Cedars-Sinai Medical Center

TABLE OF CONTENTS

PROTOCOL SUMMARY

Purpose and Knowledge to be Gained	• The purpose of the research is to study the precision and accuracy of artificial intelligence (AI) decision support for assessing cardiac structures and function, and understand clinician interactions with AI decision support systems
Research Procedures	• The primary research procedures are prospective, blinded, randomized trial of AI decision support in the echocardiography lab with review of previously acquired medical imaging.
Subject Population	• Adult patients who underw**ent** echocardiography at CSMC and/or MDRH • CSMC and/or MDRH healthcare workers caring for patients identified with ADRD
Duration	• The total study duration is 24 months.

GENERAL INFORMATION

CSMC Co-Investigators	Susan Cheng, MD, MMSC, MPH; Director, Public Health Research and Cardiovascular Population Sciences, Smidt Heart Institute, Cedars-Sinai Medical Center Joseph Ebinger, MD, MS; Assistant Professor of Medicine, Department of Medicine, Cedars-Sinai Medical Center
Sponsor/Funder	Cedars-Sinai Medical Center

1.0 BACKGROUND, RATIONALE

Recent advances in machine learning and image processing techniques have shown that machine learning models can identify features unrecognized by human experts and more precisely/accurately assess common measurements made in clinical practice. In echocardiography, this ability for precision measurement and detection is important in both disease screening as well as diagnosis of cardiovascular disease.

Echocardiography is routinely and frequently used for diagnosis and prognostication in routine clinical care, however there is often subjectivity in interpretation, heterogeneity in application, and variance with image acquisition and quality. The cardiac function, as described by the left ventricular ejection fraction (LVEF), is the focal measurement of echocardiography, and used to diagnose heart failure and determine various interventions and medical treatments. In preliminary work, we developed a novel AI algorithm to assess LVEF and showed it to be more precise

than human interpretation in 10,030 echocardiograms done at Stanford University (Ouyang et al. Nature, 2020).
Echocardiography is the perfect place to study the impact of AI models, given the step-wise, time/location staggered evaluation by multiple independent clinical experts. Currently, sonographer technicians 1) acquire ultrasound images at the bedside and 2) provide preliminary interpretations prior to validation and overreading by cardiologists. This staggered, stepwise evaluation allows for the introduction of AI decision support at the time of the sonographer's preliminary interpretation and evaluation of which the cardiologist ultimately prefers.

As part of our previously published study, at Stanford University we prospectively evaluated our model output in 51 patients showing higher precision than human readers and in blinded testing showed found clinician experts to think of AI decision support to be equivalent to sonographer preliminary readings in 40 additional patients.

To determine whether an integrated AI decision support can save sonographer time and improve accuracy of cardiologist's assessment of echocardiograms, we will conduct a blinded, randomized controlled study to test the feasibility, acceptability and effectiveness of AI guided measurements compared to sonographer guided measurements in preliminary readings of echocardiograms in the Cedars-Sinai Health System.

2.0 STUDY OBJECTIVES

Aim 1. To examine the feasibility of AI assessment of identifying apical-4-chamber view videos, identifying systole and diastole, and measuring the left ventricle to calculate end-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF).

Aim 2. To evaluate the acceptability and effectiveness to cardiologists of the AI assessments of EDV, ESV, EF in comparison with sonographer assessments of EDV, ESV, EF in a blinded fashion.

Aim 3. To examine the feasibility of AI assessment of identifying parasternal long axis view videos, measuring the left ventricle to calculate end-diastolic intraventricular septal (IVS) width, left ventricular internal diameter (LVID), and posterior wall thickness (PWT), and screen for cardiac diseases such as cardiac amyloidosis and hypertrophic cardiomyopathy based on abnormal measurements and features.

Aim 4. To examine the feasibility, acceptability, and effectiveness of AI disease screening model in identifying patients with cardiac diseases such as cardiac amyloidosis and hypertrophic cardiomyopathy, with follow-up and evaluation by clinicians based on AI screening criteria.

3.0 STUDY POPULATION

3.1 SELECTION OF THE STUDY POPULATION

The study population will include adult patients who underwent imaging in the echocardiography/non-invasive cardiac imaging laboratory.

3.2 INCLUSION CRITERIA

The study population will include patients who underwent imaging (limited or comprehensive transthoracic echocardiogram studies) in the echocardiography/non-invasive cardiac imaging laboratory.

3.3 EXCLUSION CRITERIA

We will exclude transesophageal echocardiogram studies as well as pregnant women

3.4 SUBJECT SCREENING AND ENROLLMENT

Participants will be identified by the imaging database system (Siemens Syngo). Associated subject records will also be reviewed by comprehensive chart review by a trained research coordinator. The minimum number of variables needed include: age, gender, race/ethnicity, diagnoses (using ICD 10 diagnosis codes, keywords in past medical history, problem list, clinical notes), and the medication list).

3.5 SUBJECT RECRUITMENT

For healthcare workers/providers: Sonographer preliminary tracings and interpretations will be stored and compared with both AI decision support evaluations and final physician interpretations and reports. Imaging will be identified by the imaging database system (Siemens Syngo) of previously reported echocardiographic studies.

Study participants (healthcare workers/physicians) will be approached during faculty meeting to participate in the study. An information sheet will be presented to them upon initial contact. Reminder emails will be sent weekly while study is running or until the healthcare worker/provider declines to participate.

4.0 STUDY DESIGN AND METHODS

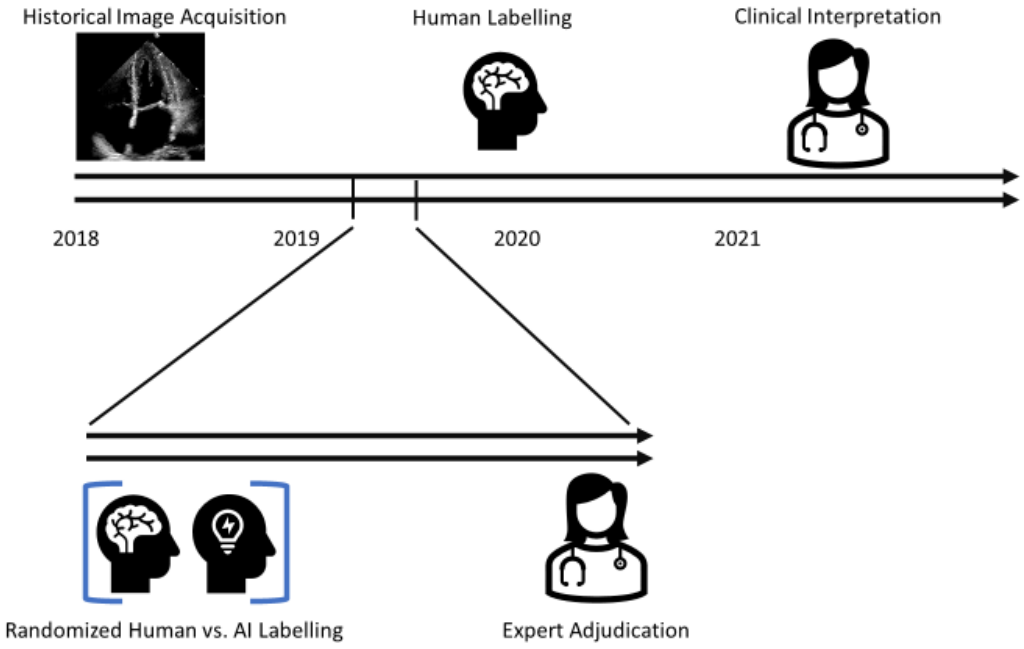
Standard Clinical Workflow

Echocardiography is the perfect place to study the impact of AI models, given the stepwise, time/location isolated evaluation by multiple independent clinical experts. The current, standard of care, clinical workflow involves three independent steps:

1. First, sonographers acquire ultrasound images from the ultrasound machine. The images as well as linear measurements then migrates from the ultrasound machine to a server for subsequent interpretation.
2. Second, sonographers provide a preliminary interpretation that includes measurements, calculations, and categorial interpretations. This is done at the computer and isolated in time from the initial image acquisition.
3. Third, cardiologists review the preliminary interpretation, making changes as necessary, and produces a final interpretation that is sent to the electronic medical record.

Currently, sonographer technicians 1) acquire ultrasound images at the bedside and 2) provide preliminary interpretations prior to validation and overreading by cardiologists. This staggered, stepwise evaluation allows for the introduction of AI decision support with minimal impact on patient care. Physicians are already used to adjusting the preliminary report given the variable training of sonographers and on the lookout for changes, variation, or adjustments that need to be made. It is common to have significant changes from the preliminary report and the preliminary report findings, and preliminary findings never interact with the rest of the healthcare system unless explicitly reviewed and confirmed by a cardiologist.

Embedded Study Workflow



Working with the image reporting system vendor, we've developed a blinded, automated process to change labels and storing initial sonographer tracings. Our workflow does not change the clinical workflow for cardiologists – they simply see a preliminary interpretation and/or measurement and are given the opportunity to change, adjust, correct, or keep the same as needed. Similar software from multiple vendors are routinely used and already built into the clinical workflow. As part of the standard of care, the CSMC echo lab currently uses Philips QLab as well as TomTec AutoLV, however they are being used haphazardly and without randomization. Randomization and blinding allows for clearer understanding of the impact of the software, but both the use of software aids, not using software aids are within the current standard of care, as well as adjudication without any kind of preliminary report.

With randomization, a proportion of the preliminary interpretations will be done by AI technology and the study team will assess how different this preliminary interpretation is from the final interpretation. Studies will be randomized 1:1 to either sonographer preliminary report finding or AI preliminary report finding with final adjudication by the cardiologist. The initial plan is for 1:1 allocation of randomization, with each study only having 1 preliminary report.

In this study, we will prompt sonographers and cardiologists to re-review historically collected and reported studies. There is no interaction with patients and no changes will

be made to the clinical report (no change to standard of care), but a limited dataset of previously acquired echocardiogram studies will under blinded re-review by sonographers and cardiologists. Either the historical clinical tracing obtained by the sonographer previously or a new sonographer tracing will be used as the human comparison. The AI comparison will be the AI algorithm's output on the previously reported images. In blinded fashion, cardiologists will re-review clinical measurements and finalize an interpretation which will be reviewed with initial preliminary report (either by sonographer or AI) as well as original clinical interpretation as a metric of retest variation. At the end of the study, participants will be asked whether the blinding and randomization was successful. Questions will be targeted to seeing if participants could identify which preliminary reports were made by software.

Cardiologists can and often do make interpretations even without a preliminary interpretation, and often the preliminary interpretation by sonographers vary greatly from the eventual cardiologist interpretation. The study is no greater than minimal risk and will have no direct impact on patient's rights, welfare, or clinical care. There is no direct interaction with patients.

We will notify sonographers and cardiologists about the use of randomization and blinding and an information sheet will be provided. One information sheet will be provided to both sonographers and cardiologists that delineate their roles.

5.0 DATA COLLECTION AND MANAGEMENT

5.1 DATA PROCUREMENT

- Data will be collected by the investigator and study staff directly from the electronic medical record and cardiology clinical imaging system. Procedures for this study will not require interaction or interventions with participants.
- A Waiver of Consent/HIPAA Authorization will be sought from the IRB because the requirement to obtain consent would make the conduct of the research impracticable due to the following reasons:
    - Given the patient population and date range of the study, subjects would likely be lost to follow-up or deceased. The proportion of individuals likely to have relocated or died may be a significant percentage of the subject population and the research results may not be meaningful and lose statistical power.
    - Contacting patients would require recording of identifiable information, such as phone numbers and addresses, beyond what is required to conduct the study.
    - As this is a population-based study, scientific validity would be compromised if consent was required. Including only those samples/records/data for which consent can be obtained would prohibit conclusions to be drawn or bias the sample such that conclusions would be skewed.

- The study is no greater than minimal risk and will have no direct impact on patient's rights, welfare, or clinical care. There is no direct interaction with subjects. The research results will not impact subjects' clinical care.
- Measures described in the Confidentiality section below will be implemented to minimize risk of a breach of confidentiality during record review and data collection/analysis.

5.2 TIME PERIOD OF DATA UNDER REVIEW

- Imaging data from 1/1/2010 to 9/1/2021 will be reviewed. Interpetations provided by sonographers and physicians will be reviewed from between the dates 09/01/2021 to 10/30/2023.
- All research data and related information will be retained in a secure access-controlled location for at least 2 years and up to 7 years after completion of the study.

5.3 VARIABLES COLLECTED

- Imaging data, including measurements, interpretations, and calculations from the clinical imaging system. From the electronic medical record system, patient demographics and diagnoses.

5.4 SOURCE DOCUMENTS

- All clinical data will be extracted from the EHR and clinical imaging database as described above.

5.5 DATA COLLECTION AND STORAGE

- **Secure storage**: Data will be collected via EHR or clinical imaging database. All data will be housed in a HIPAA-compliant secure storage system, like REDCap or Box, within the Cedars-Sinai network with access restricted to approved members of the research team.

- **Limited Access:** Private identifiable information, will be accessible only to IRB approved study team members with current IRB training.

- **Unique ID Numbers:** Each patient will be assigned a unique ID number, which will be used to code data and specimens.

- **Removal of Identifiers:** Direct identifiers (e.g. name or MRN) will be removed from any research records and destroyed as soon as scientifically possible and maintained only as long as necessary to abstract, analyze and verify data. Only de-identified coded data will be available for approved data analysts to perform statistical analyses.

- **Storage of Physical Records:** Any physical records for this study will be maintained at a secure location where access is limited to approved personnel. The records will not be removed from Cedars-Sinai premises.

5.6 CONFIDENTIALITY AND SECURITY OF DATA

To protect confidentiality, all data will be stored in a password-protected computer within CSMC firewall. In addition, information will be input into a secure, password protected database, to which only a select group of research personnel will have access.

6.0 DATA AND SAFETY MONITORING

6.1 DATA AND SAFETY MONITORING PLAN

N/A - The study is no greater than minimal risk and will have no direct impact on patient's rights, welfare, or clinical care. There is no direct interaction with subjects. The research results will not impact subjects' clinical care.

6.2 QUALITY CONTROL AND QUALITY ASSURANCE

Data quality will be overseen directly by the Principal Investigator. To ensure overall data integrity, a data analyst will review the imaging database weekly and confirm the data points have been correctly identified and recorded. Discrepancies will be brought to the attention of the PI. The data analyst will record his/her activities and findings weekly in a report filed with the PI, which will be made available to the IRB at the end of the study period, or at any time upon request

7.0 STATISTICAL CONSIDERATIONS

7.1 STUDY OUTCOME MEASURES

The primary outcome measures are:

1. Effects of the AI systems decision in improving clinician efficacy and efficiency:
    a. Decrease in the time to complete each imaging study
    b. Decrease in the frequency and amount of change from the preliminary report to the final report
    c. Evaluation of distribution of interpretations in cardiologist's final interpretations

2. Effects of the AI systems integration with satisfaction, computer-human interaction

a. Evaluation of whether cardiologists can distinguish between AI vs. sonographer preliminary measurements.
b. Evaluation for potential changes in human behavior over time.

7.2 SAMPLE SIZE CONSIDERATIONS

Sonographers are expert adjudicators who create preliminary reports before review by the cardiologist. While there might be small disagreements within measurement error, we define meaningful change as changes beyond 5% of a metric (IE, going from 45 to 50% in the ejection fraction) or a change in classification (going from mild to moderate). In this setting, we estimate metrics and measurements are meaningfully changed from the preliminary sonographer report by the cardiologist **8% of the time.** We hope to find a statistically meaningful decrease in the frequency of change when using AI decision support, such that the preliminary AI report is changed by the cardiologist **5% of the time.**

The Cedars-Sinai Non-Invasive (Echocardiography) Laboratory is staffed by 10 cardiologists and ~approximately 50 sonographers, who will participant in the study.

With an alpha of 0.05, power of 0.9, and a 1:1 enrollment ratio, we anticipate needing a sample size of 2834 studies, with 1417 studies in each arm. As a buffer against dropout and other issues, we anticipate enrolling 1750 patients in each arm for 3500 total.

8.0 REFERENCES

1. Ouyang D, Zou JY. “Deep learning models to detect hidden clinical correlates.” *The Lancet Digital Health. July 2, 2020*
2. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, Zou JY* “Interpretable artificial intelligence for beat-to-beat cardiac function assessment” *Nature. Mar 25, 2020.*
3. Ouyang D, Gorbhani A, Abid A, He B, Chen JH, Harrington RA, Liang DH, Ashley EA, Zou JY. “Deep Learning Interpretation of Echocardiograms.” *npj Digital Medicine. January 24, 2020.*
4. Augusto, J. B. *et al.* Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *Lancet Digit Health* **3**, e20–e28 (2021).

5. Howard, J. P. *et al.* Automated Left Ventricular Dimension Assessment Using Artificial Intelligence Developed and Validated by a UK-Wide Collaborative. *Circ. Cardiovasc. Imaging* **14**, e011951 (2021).
6. Zhang, J. *et al.* Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation* **138**, 1623–1635 (2018).
7. Madani, A., Arnaout, R., Mofrad, M. & Arnaout, R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* **1**, (2018).