Diagnostic Accuracy of a Handheld Ultrasound vs a Cart-Based Model: A Randomized Clinical Trial

NCT05196776

2 September 2023

Study Protocol

Study investigators utilized permuted-block randomization with an allocation ratio of 1:1. Allocation concealment included sequentially numbered, opaque, sealed envelopes. Upon enrollment, blinded study investigators selected an envelope containing study materials and pre-randomized selection into the handheld device (HH) or cart-based model (CB) using Research Randomizer (Version 4.0).[26] Patients, who required a cardiac, lung, renal, aorta, or biliary POCUS, were randomized to a portable device, the Butterfly iQ (Guilford, CT) transducer connected to a 5th generation Apple iPad Mini (Cupertino, CA), or to a cart-based model, the GE Venue Go or GE Logiq E (Wauwatosa, WI). A PGY 1-3 emergency medicine (EM) resident performed each POCUS. An EP, credentialed in the core American College of Emergency Physicians POCUS applications, reviewed each study concurrently. Residents completed all POCUS scans prior to advanced imaging.

A cardiologist-interpreted echocardiogram, performed within 24 hours of presentation to the emergency department, served as the reference standard for cardiac images. For biliary tract images, the reference standard was a radiology-interpreted ultrasound, performed during the ED visit. For lung, renal and aortic scans, the reference standard was computed tomography images (when available and performed during the ED visit), or POCUS quality assurance review by two ultrasound fellowship trained physicians (when no CT was available)." If there was disagreement, a third ultrasound fellowship trained physician provided an interpretation. The cardiologist, radiologist, and EPs were blinded to POCUS reads.

Prior to starting their internship, our emergency medicine residents participate in an introductory five-hour introduction to POCUS course taught by our emergency ultrasound faculty. Additionally, each resident completes a three-week emergency ultrasound rotation

during their internship in accordance with Accreditation Council for Graduate Medical Education (ACGME) and ACEP guidelines.[24,27] No additional training was provided to the residents prior to their study participation. Nonetheless, each participant completed more than 25 of each scan prior to participating in the study to achieve competency per ACEP and AGME guidelines.[24,27]

Measurements

Prior to study commencement, study investigators defined the following diagnostic endpoints: ejection fraction (good >50%, moderate 30-50%, poor <30%) and the presence or absence of the following: gallstones, hydronephrosis (mild, moderate, or severe), B-lines ($\geq$3 in a single lung field), and abdominal aortic aneurysm (>3cm).

Using the electronic medical record, Epic (Verona, WI), study investigators performed chart abstraction on all patients to obtain results of cardiology-interpreted echocardiograms and radiology-interpreted ultrasound and computed tomography studies.

The primary endpoint was diagnostic accuracy of each imaging modality compared to the aforementioned gold standards. The secondary endpoint of image quality was assessed by three ultrasound fellowship-trained physicians using a previously validated Likert scale assessing image quality.[28] A score of 1 indicated unable to interpret and a score of 7 specified superior imaging quality.

Statistical Analysis

Given the lack of pre-existing data comparing the modalities, investigators hypothesized that the cart-based model would have a 90% diagnostic accuracy versus 70% for the handheld device. We postulated that the HH would be inferior given the smaller screen size, novel technology to generate sonographic images, and limited provider experience with the device. To

detect a 20% difference, the sample size calculation of 98, with 49 patients randomized to each arm, was based on a power of 80% with an alpha of 0.05. We report continuous and categorical data as medians with interquartile ranges (IQR) or proportions with 95% confidence intervals (CIs) and utilized standard 2x2 tables to calculate test characteristics with 95% confidence intervals (CIs). Intraclass correlation coefficient (ICC) assessed inter-rater reliability between blinded expert reviewers, and the t-test compared median Likert scores. All analyses were performed using MedCalc (Version 19.1.6).