OPEN

Original Research

Virtual Compared With In-Clinic Transvaginal Ultrasonography for Ovarian Reserve Assessment

Esther H. Chung, MD, Laura C. Petishnok, MS, Jesse M. Conyers, MD, David A. Schimer, MD, Wendy S. Vitek, MD, Amy L. Harris, MD, Michelle A. Brown, MD, Julie A. Jolin, MD, Anatte Karmon, MD, and Aaron K. Styer, MD

OBJECTIVE: To evaluate noninferiority of virtual transvaginal ultrasonography compared with in-clinic ultrasonography for ovarian reserve assessment.

METHODS: We conducted a single-site, head-to-head crossover trial. Participants performed self-administered virtual transvaginal ultrasonography at home, guided by a remote-certified ultrasound technologist, then underwent transvaginal ultrasonography in-clinic with another ultrasound technologist. Participants were women in the greater Boston area interested in evaluating ovarian reserve and recruited through social media, health care referrals, and professional networks. The uterus and ovaries were captured in sagittal and transverse views. These randomized recordings were reviewed by two or three independent, blinded reproductive endocrinologists. The primary outcome was noninferiority of the rate of clinical quality imaging produced at home compared with in clinic. Sample size was selected for greater than 90% power, given the 18% noninferiority margin. Secondary outcomes included antral follicle count equivalency and net promoter score superiority.

RESULTS: Fifty-six women were enrolled from December 2020 to May 2021. Participants varied in age (19–35 years), BMI (19.5–33.9), and occupation. Ninety-six percent of virtual and 98% of in-clinic images met "clinical quality." The difference of −2.4% (97.5% CI lower bound −5.5%) was within the noninferiority margin (18%). Antral follicle counts were equivalent across settings, with a difference in follicles (0.23, 95% CI −0.36 to 0.82) within the equivalence margin (2.65). Virtual examinations had superior net promoter scores (58.1 points, 97.5% CI of difference 37.3–79.0, $P<.01$), indicating greater satisfaction with the virtual experience.

CONCLUSION: Virtual transvaginal ultrasonography remotely guided by an ultrasonography technologist is noninferior to in-clinic transvaginal ultrasonography for producing clinical quality images and is equivalent for estimating antral follicle count. Virtual transvaginal ultrasonography had superior patient satisfaction and has potential to significantly expand patient access to fertility care.

FUNDING SOURCE: This study was sponsored by Turtle Health.

CLINICAL TRIAL REGISTRATION: ClinicalTrials.gov, NCT04687189.

(Obstet Gynecol 2022;139:561–70)

DOI: 10.1097/AOG.0000000000004698

From the Duke University Medical Center, Durham, North Carolina; Boston Children's Hospital, Boston, Massachusetts; Emory University, Atlanta, Georgia; University of Rochester Medical Center, Rochester, New York; Wright State University, Wright-Patterson AFB, Ohio; Northwestern Medicine, Chicago, Illinois; Massachusetts General Hospital, Boston, Massachusetts; Hawaii Pacific Health, Honolulu, Hawaii; and CCRM Fertility, Boston, Massachusetts.

This study was sponsored by Turtle Health.

Presented at the American Society for Reproductive Medicine's Scientific Congress and Expo, October 16–20, 2021, Baltimore, Maryland.

The authors thank Jessica Galante, Lauren Cristoferi, Deborah A. Hubacz, Kerileigh Dorion, and Erin Silliman, MS, for their contribution to this work.

Each author has confirmed compliance with the journal's requirements for authorship.

Corresponding author: Esther H. Chung, MD, Duke University Hospital, Durham, NC, 27710; Email: esther.h.chung@duke.edu.

Financial Disclosure:

The authors Esther H. Chung, Laura C. Petishnok, Jesse M. Conyers, David A. Schimer, Wendy S. Vitek, Amy L. Harris, Michelle A. Brown, Julie A. Jolin, Anatte Karmon and Aaron K. Styer are paid consultants and advisors of Turtle Health. None hold equity or options in the Company. The other authors did not report any potential conflicts of interest.

ISSN: 0029-7844/22

Infertility affects one in eight reproductive-aged couples in the United States.[1] However, diagnostic testing for this disorder is often not pursued due to cost, lack of insurance coverage, and underestimation of age-related fertility decline among women of reproductive age.[2] Recently, with increasing average maternal age and more public discussion of infertility, there is expanding interest in at-home fertility testing as a cost-considerate option that is easy to access.[3,4] Early fertility-assessment products focused on estimating ovulation, whereas newer tests have attempted to more directly evaluate ovarian reserve.[5,6]

Ovarian reserve, the quantity and quality of oocytes, declines with advancing age and directly affects reproductive potential.[7,8] Typically, ovarian reserve is estimated with both ultrasonography to assess antral follicle count and serum antimüllerian hormone.[9] Antral follicle count is the most accurate imaging modality to estimate ovarian reserve[8–13] and is established as an essential part of a complete fertility assessment.[4,8,13] Current home-testing products are limited to fingerstick testing measuring reproductive hormones (eg, antimüllerian hormone, follicle stimulating hormone, estradiol), but they do not offer clinically meaningful ovary-specific testing as with antral follicle count.[14,15]

Virtual ovarian reserve testing with self-administered transvaginal ultrasonography to evaluate antral follicle count and pelvic organs has the potential to improve access to fertility testing, identify diminished ovarian reserve[9,13,16] earlier, and guide decisions for family building or elective fertility preservation.[15] This study evaluated the ability of virtual transvaginal ultrasonography to consistently produce clinical quality images sufficient to evaluate ovarian reserve. Specifically, the study primarily assessed whether the rate of clinically interpretable images was noninferior to that of traditional in-clinic ultrasonography and whether the resulting estimates of antral follicle count were equivalent. This study, and associated regulatory filing, are currently under review by the U.S. Food and Drug Administration (FDA).

METHODS

The SELF-HELP validation study (Sonograms Enable Looking Forward - Home Examinations Led by Providers) was conducted as a prospective, head-to-head crossover noninferiority clinical trial in Boston, Massachusetts, and study design was reviewed by the FDA.[17] Institutional research ethics board approval was obtained from IntegReview IRB,[18] a third-party institutional review board accredited by the Association for the Accreditation of Human Research Protection Programs. The trial was registered on ClinicalTrials.gov as NCT04687189.[19]

The study was conducted as described in the institutional review board–approved study protocol, including the changes made in Major Amendment 1, which stemmed from challenges recruiting patients with known submucosal leiomyomas and FDA feedback. Two study arms were planned: 1) 30 healthy patients and 2) 15 patients with a history of submucosal leiomyomas, with institutional review board–approved recruitment materials designed for each population. Despite 3 months of recruitment, no patients with known submucosal leiomyomas were identified. As a result, the leiomyoma study arm was eliminated. Additionally, the methods were updated to address the FDA's feedback that the statistical independence of organs needed to be tested for intraparticipant correlation. Finally, in response to the changes, the total sample size of the healthy trial arm was increased to a target of 55, based on the number of participants required to fully power the primary endpoint and projected capacity of the trial site. Power calculations remained robust, and the recruitment materials were unaffected. This amendment was submitted before the review of any cine clips by the independent raters; thus, the planning of analysis remained a priori. Primary endpoints were assessed with a paired noninferiority test using a restricted maximum likelihood estimation–based test statistic.[20]

Clarius Mobile Health is a health care company based in British Columbia. The Clarius Ultrasound Scanner (K192107) is a cordless, wireless, application-based, portable transvaginal ultrasound system for medical professionals, with an existing 510(k) FDA clearance for in-clinic use.[21] It operates at a frequency of 3–10 MHz. The trial used a base setting of 8 cm for depth. The probe size (328×78×38 mm) and weight (410 g) allow for single-handed use. Each probe (model EC7 HD) was paired with an iPhone X with the Clarius Ultrasound application installed. The Clarius Live feature was enabled, allowing for "tele-ultrasound imaging." The probe was used under the "IVF" application setting without modification and as described by the manufacturer's instructions, with the exception of the novel user (the patient) and setting (home).

Participants were recruited by physicians and through social media, email listservs, and professional networks. Eligibility criteria included women between 18 and 38 years of age (inclusive), body mass index (BMI, calculated as weight in kilograms divided by height in meters squared) up to 40, and the ability to independently give consent electronically. Electronic

consent required that a participant speak native or fluent English and have a high school degree or equivalent. The study protocol authorized the principal investigator to use their professional judgment if there were a question about patient ability to give informed consent virtually, but, in practice, all patients meeting study criteria were invited to consent. All participants were required to live within a 200-mile radius of Boston, including states covered by the principal investigator's medical license during the coronavirus disease 2019 (COVID-19) emergency. Criteria were kept intentionally broad to reflect the scope of women who might be interested in participating, excluding only for specific medical or regulatory concerns. These exclusion criteria included health care professionals with expert ultrasound experience (eg, ultrasound technologists, radiologists, obstetrician–gynecologists), those with BMIs higher than 40, those unable to speak English fluently, those who were pregnant or breastfeeding, and those at fewer than three postpartum menstrual cycles from delivery or fewer than two cycles after miscarriage.

Eighty potential participants were initially screened. Of those, 64 had virtual consent appointments, signed electronic informed consents, and submitted prestudy questionnaires that gathered relevant clinical information, including age, birthdate, height and weight, menstrual cycle regularity, last menstrual period (if known), obstetric history, any prior professional ultrasound experience, any prior transvaginal ultrasonography in clinic, and current contraception. Of those consented, 56 participants were ultimately enrolled. Of the other eight, four stopped responding, one left the trial catchment area, one withdrew for unspecified reasons, one was unable to schedule, and one disclosed disqualifying medical information after consent. To reduce the risk of inadvertent fetal exposure, a regulatory concern to the FDA,[22] women with regular menses were scheduled between days 3 and 10 of their menstrual cycle. Women with oligomenorrhea or amenorrhea were scheduled on consecutive days as well but at any time in their cycle.

Before the examination, participants were shipped a Turtle Health package that included a Clarius transvaginal ultrasound probe, an iPhone, two lubricant packets, two probe covers, chargers for both probe and iPhone, instructions, and a Turtle logo sticker. During the virtual visit, the participant performed a self-administered virtual transvaginal ultrasound examination with telemedicine guidance by an American Registry for Diagnostic Medical Sonography–certified ultrasound technologist. Before the study, these ultrasound technologists underwent a proprietary training program, including study materials, simulated exercises, and a separate precursor trial with live patient examinations. Under the ultrasound technologist's direction, six cine clips were taken for each participant: sagittal and transverse views of the uterus, right ovary, and left ovary. The ultrasound technologists used key commands to guide the participant through the examination, including: "insert the probe, "go slower," "rotate the probe 90° to your left," "move your hand to your right or left," "press down [if gas seen]," and "move the probe up or down."

The next day, the participant brought the Turtle Health package to a clinical site where a different ultrasound technologist of similar experience level and training conducted the in-clinic transvaginal ultrasonography examination with the same probe. Ultrasound technologists switched roles throughout the study so that the in-person and virtual roles were approximately evenly distributed. Any given participant had a different ultrasound technologist for the virtual and in-person scans. Additional precautions were taken to minimize bias, including instructing the ultrasound technologists not to communicate with each other about the participants.

All cine clips were uploaded to a Health Insurance Portability and Accountability Act–compliant Google Drive folder and randomized numerically. Two independent, blinded raters (board-certified reproductive endocrinologists) reviewed the videos, recorded their ratings for clinical quality, and quantified total antral follicle count, if possible. A third reproductive endocrinologist was available as a tie-breaker in the event of discrepant ratings on clinical quality.

All participants who completed the study received their results and were compensated for participation at fair market value. Any participant with new pathology on their in-person scans, identified by both raters on their surveys with high certainty, met with the principal investigator directly to review results. Confirmatory in-clinic scans, independent of the experimental process, were recommended as indicated.

Cine clips for each organ captured at home and in clinic were assigned to folders named based on an Excel-generated list of 500 four-digit random integers using 500 instances of the formula ROUND(10,000*RAND(),0). Each folder contained two views (sagittal and transverse) of a single organ (left ovary, right ovary, or uterus) obtained from a single participant in a single setting (at home virtual

or in person). There were 334 such folders in total (56 participants×3 organs×2 modalities−2 folders owing to a participant with only one ovary). Folders were sorted based on their file number from smallest to largest within each batch of videos provided to the raters, randomizing the order they were viewed by the raters. Video folders were provided to raters in batches that included, at minimum, all cine clips (three organs in two different settings) from six different participants, totaling a minimum of 36 folders per batch. Raters were blinded to the organ, patient, and setting and viewed a random sequence of organs from multiple participants and settings in each batch.

Given lack of an existing consensus for transvaginal ultrasound image quality criteria,[23–25] a 4-point scale was developed from interviews with expert physicians who regularly assess images. During scale development, emphasis was placed on replicating and codifying the thought process from usual clinical practice and ensuring that all clips could be intuitively categorized by an expert physician. The scales for imaging quality assessment used by Hausleiter et al[26] were used as a guide for quality rating.

For each folder, the raters were asked to score the visualization of the organ on a defined quality scale from 1 (nondiagnostic) to 4 (excellent image quality) (Appendix 1, available online at http://links.lww.com/AOG/C594).[26] For the purposes of assessing the primary endpoint, a "clinical quality" scan of an organ required two raters to provide a score of either 3 or 4 and that the same raters identified the target organ correctly (ie, complete visualization of the ovary [cortex, stroma, antral follicles] in perpendicular planes [sagittal, transverse]). If the rater did not accurately identify the target organ, these scans were considered beneath clinical quality and the folder was rated as less than the threshold for clinical quality (scored as 1).

Each primary rater provided ratings for all 334 such folders. The third rater rated 15 of 334 (4%) folders (eight at-home scans, seven in-clinic scans) when there was a disagreement on quality between the primary raters. If the tiebreaker assessed the clips to be clinical quality, this third rater's responses were used in place of those rating the images as below clinical quality.

For each folder, the raters also provided an estimated total antral follicle count. They indicated the presence or absence of submucosal leiomyomas, denoted any other observed pathology, and indicated whether clinical follow-up was recommended.

The primary study endpoint was cine clip clinical quality for each organ imaged. Secondary endpoints included equivalency of assessing antral follicle counts and superiority of virtual compared with at-home net promoter score. Net promoter score is an established, validated metric of customer experience in a variety of commercial and clinical domains[27–32] and is typically used to assess preferences between similar services or reactions to potential changes in a service or product. Other secondary endpoints included submucosal leiomyoma specificity and equivalency of identifying other nonreplicable findings (eg, cysts).

Sample size was selected to ensure greater than 90% power to demonstrate noninferiority of the clinical quality image rate, given a noninferiority margin of 18%, a true margin of up to 7%, a 95% clinical quality image rate in-clinic, an adjustment for up to a 60% intrapatient correlation between organ images, and an alpha of 2.5%. This alpha was chosen based on convention for one-sided tests, using FDA guidance.[17] Antral follicle count equivalency was well-powered (greater than 90%) under reasonable assumptions.

Power calculations for sample size and CIs were based on a noninferiority test for paired binary data using a restricted maximum likelihood estimation–based test statistic.[17,20] The noninferiority margin was chosen using FDA guidance that the maximum acceptable margin (18%) can be justified in cases of major improvements in tolerability, access to care, and lack of irreversible outcomes.[17,33] In practice, such a generous margin proved unnecessary–the trial was still more than 85% powered to demonstrate noninferiority at a 10% noninferiority margin, given the observed delta of 2.4%.

The primary endpoint analysis treated each organ from each participant as a separate data point. Importantly, independence of intraparticipant image ratings was not assumed. Multiple randomization events from the same participant were downweighted based on the intrapatient correlation in clinical image quality (0.32). This adjustment formula was defined a priori in the study protocol in response to FDA feedback to test independence of ratings. Additionally, organs obtained in at-home scans were excluded if acquired in less time than was deemed to be a credible independent process from the previous cine clip (less than 30 seconds to locate the subsequent organ). This process excluded nine organs (5% of images obtained at home), which had all been rated as clinical quality. After adjustment, each participant contributed, on average, the statistical equivalent of 2.05 independent observations to the primary endpoint across the three organs imaged. Despite their

complexity, these adjustments had little effect on the study's conclusions because each organ met the predefined noninferiority endpoint when assessed separately.

Paired continuous outcomes (antral follicle count and net promoter score) were assessed with a standard paired *t* test for equivalence and for one-sided significance.[34,35] The null hypothesis for antral follicle count estimates was that the difference between measured antral follicle counts at home and in clinic for each organ would be outside the prespecified margin of equivalence of 2.65 follicles. The 2.65 follicles represents the average intercycle variation reported in the literature.[36–38] Additional agreement testing for antral follicle count was performed, including intraclass correlation both between settings and between raters in each setting. Interrater agreement within each setting (measured by intraclass correlation) was also compared, because evidence suggests a positive correlation between interrater reliability of antral follicle count and image quality in transvaginal ultrasonography.[11]

Net promoter score is an established comparative measure of customer satisfaction or preference, as measured by the debrief survey question, "How likely are you to recommend XX to a friend or colleague?" [on a scale from 0 to 10].[28] The numerical score is a result of subtracting the percentage of detractors (scores of 6 or lower) from the percentage of promoters (scores of 9–10). Outcomes can range from −100 to +100 and indicate overall satisfaction with the experience.[27–30] The value of this measure lies in its ability to compare two similar options, with the higher net promoter score likely translating to significantly better word-of-mouth referrals for a certain product, experience, or company over another in the same arena. Any net promoter score from −100 to 0 would warrant improvement; a net promoter score between 30 and 100 would be deemed a positive to great experience.[29] The net promoter score for virtual scans was expected to be significantly higher than the net promoter score for recalled in-clinic scans (for patients with previous transvaginal ultrasound experience). Patients without previous transvaginal ultrasound experience who provided a net promoter score for virtual at-home scanning were coded as a neutral or 0 net promoter score for recalled in-clinic scans. Because recalled in-clinic net promoter score was overall negative, this had the effect of improving the average recalled in-clinic net promoter score.

The other secondary outcomes tested included submucosal leiomyoma specificity and the rate of nonreplicable findings. Nonreplicable findings were cases in which a significant finding on an at-home examination was not seen during subsequent in-clinic scan. A significant finding was defined as either at least one rater noting they "definitely" saw a finding for which follow-up was recommended or two raters noting they were "almost certain" about it. Any significant finding on an at-home scan not seen on an in-clinic scan was considered nonreplicable.

Type 1 error (α) was held to 0.05 and controlled by a multiplicity adjustment across all aforementioned endpoints. This was done using a combination of sequencing and a truncated Hochberg step-up procedure. All statistical analyses were performed on the 2019 version of Microsoft Excel.

ROLE OF THE FUNDING SOURCE

The sponsor, Turtle Health, funded this study. Turtle Health was involved in the study's design, execution, data analysis, and reporting through the work of its employees and paid consultants. Turtle Health has a contract with Clarius Mobile Health, which provided the transvaginal probes for use as portable ultrasound scanners, which are FDA-cleared for in-clinic use.[39] Esther Chung, Laura Petishnok, Jesse Conyers, and Aaron Styer, paid consultants for Turtle Health, were involved in the study ideation, design, analysis, interpretation of the data, writing of the manuscript, and decision to submit the report for publication. Austin Schimer, Wendy Vitek, and Amy Harris were compensated as independent blinded raters and contributed to the manuscript only after they were unblinded after completed data analysis. Michelle Brown, Julie Jolin, and Anatte Karmon, also paid consultants, were also involved in the study ideation and design and contributed to the manuscript.

The authors had access to relevant aggregated study data and other information (such as study protocol, analytic plan and report, validated data table, and clinical study report) required to understand and report research findings. The authors take responsibility for the presentation and publication of the research findings, have been fully involved at all stages of publication and presentation development, and are willing to take public responsibility for all aspects of the work. All individuals included as authors and contributors who made substantial intellectual contributions to the research, data analysis, and publication or presentation development are listed appropriately. The role of the sponsor in the design, execution, analysis, reporting, and funding is fully disclosed. The authors' personal interests, financial or nonfinancial, relating to this research and its publication have been disclosed.[40]

RESULTS

Fifty-six participants were enrolled from December 2020 to May 2021. The trial population was diverse along a number of metrics (Table 1). The mean age of participants was 27.6 (19–35) years, and mean BMI was 23.6 (19.5–33.9); 19 participants (34%) had BMIs of 25 or higher. Occupations ranged from undergraduate or graduate students to a wide range of professional careers. Of the five (9%) participants who were verified health care professionals (eg, residents, nurses, therapists), none had prior ultrasound expertise or professional experience administering ultrasonograms. Seventy-three percent reported using contraception: oral contraceptive pills (27%), intrauterine devices (41%), vaginal ring (4%), and progestin-only pills (2%). Two participants reported physical disabilities, including visual impairment and vaginismus.

The mean duration of the virtual examination, accounting for the entirety of the time the probe was inserted into the vagina, was 13 minutes 40 seconds (SD 3.93 minutes). Duration of the in-clinic examination was 8 minutes 30 seconds (SD 2.21 minutes). The mean difference was 5 minute 10 seconds (95% CI 3.98–6.37).

All participants completed one virtual and one in-clinic examination. No reportable events met the device adverse event bar, per FDA guidance. Four participants at home (7%) and three (5%) in person reported minor discomfort or nausea. There were no patient discontinuations or dropouts after the virtual transvaginal ultrasound examination.

Table 1. Participant Characteristics (N=56)

Characteristic	Value
Age (y)	27.6 (19–35)
BMI (kg/m^2)	23.6 (19.5, 33.9)
BMI 25 or higher (overweight)	19 (34)
Contraception method	
Oral contraceptive pills	15 (27)
Intrauterine device	23 (41)
Vaginal ring	2 (4)
Progestin-only pills	1 (2)
None	15 (27)
Professional occupation	
Student (undergraduate and graduate)	29 (52)
Professional (non–health care)	11 (20)
Health care workers	5 (9)
Other	9 (16)
Recent pregnancy (within 1 y)	1 (2)
Menstrual interval 21–35 d	33 (59)

BMI, body mass index.
Data are mean (range) or n (%).

The primary study endpoint, clinical quality rate, is presented in Table 2. Ninety-six percent of the virtual at-home scans and 98% of the in-clinic scans were rated as clinical quality. The difference in rates of obtaining clinical quality cine clips between settings (virtual vs in-person) was −2.5% (97.5% CI −0.064 to 0.015, $P<.01$), adjusted for intra-user quality score correlation (Table 2). Unadjusted for intra-user independence, this margin estimate was −2.4% (97.5% CI −0.055 to 0.008, $P<.01$). In a subanalysis excluding known health care professionals, the difference in rate of obtaining clinical quality clips remained noninferior, with a difference between settings of 2.7% ($P<.01$). In another subgroup analysis examining participants with BMIs higher than 25, the difference also remained noninferior, with a difference of 5.6% ($P=.02$).

Representative images from a single participant's virtual and in-clinic cine clips can be referenced in Figure 1.

Mean antral follicle count was equivalent across both settings (Table 2, Appendix 2 [Appendix 2 is available online at http://links.lww.com/AOG/C594]). Of the 56 participants, 55 had two ovaries and one had previously had an ovary removed, resulting in 111 total ovaries imaged in both settings. Equivalency analyses were performed on the 104 ovaries that received clinical quality ratings both at-home and in-clinic. Average antral follicle count per ovary was 11.27 and 11.05 for the at-home and in-person scans, respectively. The difference (0.23, 95% CI −0.36 to 0.82) was well within the proposed equivalency margin (2.65). Agreement between settings was observed, with an intraclass correlation coefficient of 0.81 (95% CI 0.74–0.87). Interrater agreement was similar in each setting, with a difference in intraclass correlation for antral follicle count at home compared with in clinic of only −0.027.

Sent to participants within 3 days of their ultrasound visits, the debrief questionnaire containing the question, "How likely is it that you would recommend the experience to a friend or colleague? (10 = strongly agree, 0 = strongly disagree)" had a 77% response rate, which is significantly higher than typical for voluntary feedback.[41]

The virtual scans had a significantly superior net promoter score compared with recalled in-clinic scans (Table 2). Across all participants, virtual scanning's net promoter score was 58.1 points higher than recalled in-clinic scanning (97.5% CI 37.3–79.0, $P<.01$). For the virtual scans, 60% of participants were promoters (participants who rated the experience a 9–10 out of 10), 26% were neutral (scores of 7–8), and 14% were detractors (scores of 6 or lower), yielding a net promoter score

Table 2. Performance of Virtual Compared With In-Clinic Ultrasound Examination*

	Virtual Examination†	In-Clinic Examination‡	Difference Between Settings (Virtual−In-Clinic) (Estimate and CI)	*P*
Primary endpoint				
Clinical quality rate	95.8%	98.2%	−2.40% (97.5% CI −0.055 to 0.008)	*P*<.01
Secondary endpoints				
NPS (patient satisfaction)	+46.5	−11.6	58.1 (97.5% CI 37.3–79.0)	*P*<.01
Average AFC/ovary	11.27	11.04	2.3% (95% CI −0.36 to 0.82)	
Submucosal leiomyoma specificity	100%	100%		
Rate of nonreplicable findings [n (%)]				
Including simple cysts§	2 (3.6)	1 (1.8)	−1.8% (97.5% CI −0.087 to 0.052)	
Excluding simple cysts	0 (0)	1 (1.8)	+1.8% (97.5% CI −0.026 to 0.061)	

NPS, net promoter score; AFC, antral follicle count.
* N=56 participants, 167 organs imaged.
† Participant conducting ultrasonogram with ultrasonographer oversight using telemedicine.
‡ Ultrasonographer conducting ultrasonogram.
§ Per-protocol, all in-clinic findings were considered "replicable in-clinic" unless not confirmed during follow-up scan.

of +46.5. For recalled in-clinic transvaginal ultrasonography, 2% of participants were promoters, 84% were either neutral (23%) or neutral owing to no previous transvaginal ultrasound experience (61%), and 14% were detractors, resulting in a net promoter score of −11.6. Evaluating only the 17 women with prior in-clinic experience, the recalled in-clinic net promoter score was −29.4, resulting in an unadjusted net promoter score difference between settings of 75.9.

No patient was identified as having a submucosal leiomyoma, despite the raters being prompted on their survey. The difference in the rate of nonreplicable findings (virtual−in-person, including simple ovarian cysts) was −1.8% (97.5% CI −8.7% to 5.2%). This difference met the predefined endpoint for noninferiority. Seven findings were seen in both settings (ie, large or complex cysts). Three simple cysts and one polyp were seen in clinic only, and two simple cysts were seen at home only.

DISCUSSION

This study demonstrates that virtual transvaginal ultrasound evaluation of antral follicle ovarian reserve assessment performed at home by a patient is clinically noninferior to in-clinic examination. A literature search as of November 11, 2021, of comprehensive databases such as Science Direct, PubMed, and Google Scholar using the search terms "self-administered ultrasound", "self-performed ultrasound", "at-home ultrasound", or "patient performed ultrasound" AND "vaginal", along with "antral follicle" and "ovarian reserve" yielded no relevant results. No such studies have been completed for healthy patients not actively undergoing infertility treatment. Additionally, the FDA accepted a De Novo filing for the device in June of 2021, indicating no existing predicated device with this intended use.

Virtual ultrasonography produced similar clinical quality images to those obtained in in-office practice, where reported rates of optimal pelvic organ

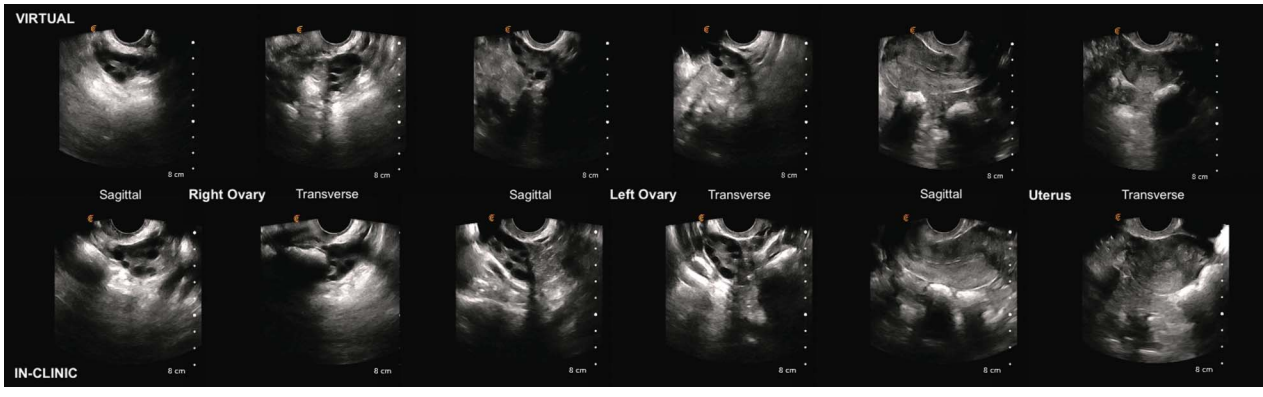

Fig. 1. Representative image panel of virtual and in-clinic cine clips of the right ovary, left ovary, and uterus.
Chung. Virtual vs in-Clinic Transvaginal Ultrasonography. Obstet Gynecol 2022.

visualization range between 90% and 99%, with average visualization rates of 92–97% seen in reproductive-aged women similar to those represented in this trial.[42–45] Although not direct comparisons, because they do not represent a population of women seeking ovarian reserve information or the device used in the trial, they do suggest that the absolute at-home imaging quality rate (96%, 97.5% CI 92–99%) is not qualitatively different than the in-office quality rates seen for other common transvaginal ultrasound use cases. These high imaging quality rates persisted after excluding participants who are health care professionals and after specifically examining participants within the highest BMI ranges. Although higher BMI was associated with slightly lower image quality rates, performance of virtual transvaginal ultrasonography in this group (93%) was consistent with prior studies using traditional transvaginal ultrasonography.[46–48]

Considering the intimate nature of transvaginal ultrasonography, patient feedback was highly favorable overall. Many patients found the virtual experience to be one they would highly recommend to others, with a net promoter score well above the health care average[28–32,49] and significantly superior to that of the in-clinic experience. As a point of reference, the net promoter score difference between at-home and in-clinic examinations (58.1 points) is nearly three times greater than the net promoter score difference between total knee replacement and total hip replacement, procedures known to have significant differences in patient tolerability.[28] The net promoter score of recalled in-clinic transvaginal ultrasound experience was, in fact, negative, potentially representing patients who would hesitate to seek care in person. Our findings suggest that this virtual testing option would be highly acceptable to patients and would improve utilization of and access to ovarian reserve testing.

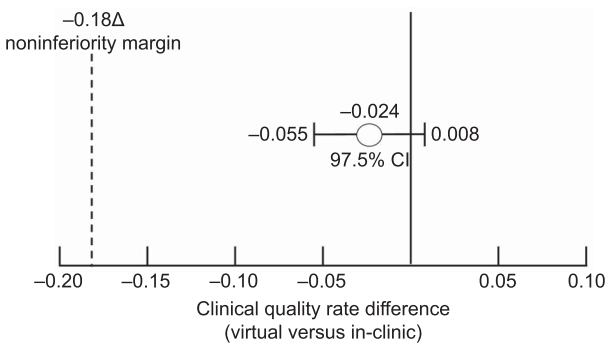

Fig. 2. Noninferiority of clinical quality rates of virtual vs in-clinic scans.

Chung. Virtual vs in-Clinic Transvaginal Ultrasonography. Obstet Gynecol 2022.

The primary strength of our study was that the population was diverse in physical characteristics (BMI), range of professions, disabilities, and contraceptive use. Hence, our results are likely generalizable to the general population of reproductive-aged women. Notably, the trial population reflected individuals with disabilities, including a participant with visual impairment and another with a history of vaginismus and prior challenges with in-clinic care. The diversity of our trial population demonstrates the broad, proactive interest of reproductive-aged persons to pursue self-evaluation of ovarian reserve.

There were several study limitations. Although the trial's participants were diverse, there were specific populations that were not represented, including patients with class II obesity or higher and non–English-speaking participants. The lack of previously validated standards for clinical quality transvaginal ultrasound imaging is a general limitation for research in this area. We hope that the endpoint used in this trial–which was found to be intuitive and representative of clinical practice by the blinded raters–may inform future studies. There was some interrater variability in antral follicle count estimation, demonstrating the inherent limitation that antral follicle count is operator-dependent. As we consider implementation of virtual ultrasonograms in clinical practice, additional limitations around scalability, including ongoing need to train ultrasound technologists, will require ongoing focus.

In the United States relative to other countries, access to high-quality ovarian reserve testing has been suboptimal and contributes to known disparities in access to infertility treatment.[4,5,50–52] Comprehensive at-home fertility assessment with FDA-authorized virtual ultrasonography (currently under FDA review) could be most effective for those who live in underserved regions, underrepresented minorities, and patients with disabilities. Because the trial demonstrated that patients can and are able to perform their own transvaginal ultrasonography, this model could possibly enable earlier identification of fertility-affecting conditions such as diminished ovarian reserve and polycystic ovarian syndrome.

This self-administered technology has the potential to reduce geographic and financial barriers to reproductive health care and provide much needed, convenient access to fertility testing to guide decisions around family building, fertility preservation, or pursuing further testing or treatment. Promising future applications may also include remote follicular monitoring for ovulation induction and controlled ovarian hyperstimulation for in vitro fertilization in tandem with treatment-specific hormonal testing. In

summary, virtual transvaginal ultrasonography can provide an innovative option for comprehensive at-home fertility testing, possibly expanding patient access to fertility care, decreasing costs, and providing women with the means to evaluate their reproductive potential.

REFERENCES

1. Centers for Disease Control and Prevention. Reproductive health: infertility FAQs. Accessed March 23, 2020. https://www.cdc.gov/reproductivehealth/infertility/index.htm
2. Lampic C, Svanberg AS, Karlström P, Tydén T. Fertility awareness, intentions concerning childbearing, and attitudes towards parenthood among female and male academics. Hum Reprod 2006;21:558–64. doi: 10.1093/HUMREP/DEI367
3. CB Insights. The body series: how technology is transforming women's healthcare. Accessed August 23, 2021. https://www.cbinsights.com/research/report/womens-health-technology-transforming/
4. Kyweluk M, Feinberg EC. Direct-to-consumer (DTC) ovarian reserve testing benefits the company, not the consumer. Accessed August 23, 2021. https://www.fertstertdialog.com/posts/59741-kyweluk-consider-this
5. Kyweluk MA. Quantifying fertility? Direct-to-consumer ovarian reserve testing and the new (in)fertility pipeline. Soc Sci Med 2020;245:112697. doi: 10.1016/J.SOCSCIMED.2019.112697
6. Yeh PT, Kennedy CE, Van der Poel S, Matsaseng T, Bernard L, Narasimhan M. Should home-based ovulation predictor kits be offered as an additional approach for fertility management for women and couples desiring pregnancy? A systematic review and meta-analysis. BMJ Glob Heal 2019;4:e001403. doi: 10.1136/BMJGH-2019-001403
7. Amanvermez R, Tosun M. An update on ovarian aging and ovarian reserve tests. Int J Fertil Steril 2016;9:411. doi: 10.22074/IJFS.2015.4591
8. Testing and interpreting measures of ovarian reserve: a committee opinion. Fertil Steril 2015;103:e9–17. doi: 10.1016/J.FERTNSTERT.2014.12.093
9. Ulrich ND, Marsh EE. Ovarian reserve testing: a review of the options, their applications, and their limitations. Clin Obstet Gynecol 2019;62:228. doi: 10.1097/GRF.0000000000000445
10. Sabek EAS, Saleh OI, Ahmed HA. Ultrasound in evaluating ovarian reserve, is it reliable? Egypt J Radiol Nucl Med 2015;46:1343–8. doi: 10.1016/J.EJRNM.2015.08.008
11. Kannamannadiar J, Campbell B, Hopkisson J, Clewes J, Johnson I, Raine-Fenning N. Establishing the intercycle variability of three-dimensional ultrasonographic predictors of ovarian reserve. Fertil Steril 2008;90:2126–32. doi: 10.1016/J.FERTNSTERT.2007.10.028
12. Fleming R, Seifer DB, Frattarelli JL, Ruman J. Assessing ovarian response: antral follicle count versus anti-Müllerian hormone. Reprod Biomed Online 2015;31:486–96. doi: 10.1016/J.RBMO.2015.06.015
13. Tal R, Seifer DB. Ovarian reserve testing: a user's guide. Am J Obstet Gynecol 2017;217:129–40. doi: 10.1016/J.AJOG.2017.02.027
14. Modern Fertility. Clinical study: reliable, easy testing for fertility hormones. Accessed August 23, 2021. https://modernfertility.com/blog/whitepaper/
15. Brezina PR, Haberl E, Wallach E. At home testing: optimizing management for the infertility physician. Fertil Steril 2011;95:1867–78. doi: 10.1016/J.FERTNSTERT.2011.01.001
16. Steiner AZ, Pritchard D, Stanczyk FZ, Kesner JS, Meadows JW, Herring AH, et al. Association between biomarkers of ovarian reserve and infertility among older women of reproductive age. JAMA 2017;318:1367–76. doi: 10.1001/JAMA.2017.14588
17. U.S. Food and Drug Administration. Non-inferiority clinical trials. Accessed August 23, 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials
18. Advarra. Accessed August 23, 2021. https://www.advarra.com/
19. ClinicalTrials.gov. Sonograms Enable Looking Forward - Home Examinations Led by Providers validation study (SELF-HELP). Accessed August 23, 2021. https://clinicaltrials.gov/ct2/show/NCT04687189
20. Liu J, Hsueh H, Hsieh E, Chen JJ. Tests for equivalence or non-inferiority for paired binary data. Stat Med 2002;21:231–45. doi: 10.1002/SIM.1012.
21. Clarius. About us.Accessed August 23, 2021. https://clarius.com/about/
22. U.S. Food and Drug Administration. Avoid fetal "keepsake" images, heartbeat monitors. Accessed November 19, 2021. https://www.fda.gov/consumers/consumer-updates/avoid-fetal-keepsake-images-heartbeat-monitors
23. Le Lous M, De Chanaud N, Bourret A, Senat MV, Colmant C, Jaury P, et al. Improving the quality of transvaginal ultrasound scan by simulation training for general practice residents. Adv Simul (Lond) 2017;2:24. doi: 10.1186/s41077-017-0056-z
24. Salomon LJ, Nassar M, Bernard JP, Ville Y, Fauconnier A. A score-based method to improve the quality of emergency gynaecological ultrasound examination. Eur J Obstet Gynecol Reprod Biol 2009;143:116–20. doi: 10.1016/J.EJOGRB.2008.12.003
25. Jin BB, Gong YZ, Ma Y, He ZH. Gynecological emergency ultrasound in daytime and at night: differences that cannot be ignored. Ther Clin Risk Manag 2018;14:1141–7. doi: 10.2147/TCRM.S169165
26. Hausleiter J, Martinoff S, Hadamitzky M, Martuscelli E, Pschierer I, Feuchtner GM, et al. Image quality and radiation exposure with a low tube voltage protocol for coronary CT angiography results of the PROTECTION II trial. JACC Cardiovasc Imaging 2010;3:1113–23. doi: 10.1016/J.JCMG.2010.08.016
27. Qualtrics. What is NPS? Your ultimate guide to net promoter score. Accessed August 23, 2021. https://www.qualtrics.com/experience-management/customer/net-promoter-score/
28. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AH, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014;96-B:622–8. doi: 10.1302/0301-620X.96B5.32434
29. Retently. What is a good net promoter score? (2021 NPS benchmark). Accessed August 23, 2021. https://www.retently.com/blog/good-net-promoter-score/
30. Krol M, de Boer D, Delnoij D, Rademakers J. The Net Promoter Score–an asset to patient experience surveys? Health Expect 2015;18:3099–109. doi: 10.1111/HEX.12297
31. Reichheld FF. The one number you need to grow. Harv Bus Rev December 81:46–54.
32. Koladycz R, Fernandez G, Gray K, Marriott H. The net promoter score (NPS) for insight into client experiences in sexual and reproductive health clinics. Glob Heal Sci Pract 2018;6:413–24. doi: 10.9745/GHSP-D-18-00068

33. Ahn S, Park S, Lee K. How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research. Radiology 2013;267:328–38. doi: 10.1148/RADIOL.12120725

34. Mara M, Horak P, Kubinova K, Dundr P, Belsan T, Kuzel D. Hysteroscopy after uterine fibroid embolization: evaluation of intrauterine findings in 127 patients. J Obstet Gynaecol Res 2012;38:823–31. doi: 10.1111/J.1447-0756.2011.01782.X

35. DJS. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 1987;15:657–80. doi: 10.1007/BF01068419

36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8. doi: 10.1037/0033-2909.86.2.420

37. van Disseldorp J, Lambalk CB, Kwee J, Looman CW, Eijkemans MJ, Fauser BC, et al. Comparison of inter- and intra-cycle variability of anti-Müllerian hormone and antral follicle counts. Hum Reprod 2010;25:221–7. doi: 10.1093/HUMREP/DEP366

38. Subirá J, Alberola-Rubio J, Núñez M, Escrivá AM, Pellicer A, Montañana V, et al. Inter-cycle and inter-observer variability of the antral follicle count in routine clinical practice. Gynecol Endocrinol 2017;33:515–18. doi: 10.1080/09513590.2017.1291614

39. K192107-Prithul Bom. Accessed August 23, 2021. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192107

40. Instructions for authors–July 2021. Obstet Gynecol 2021;138:138–55. doi: 10.1097/AOG.0000000000003955

41. Baruch Y, Holtom BC. Survey response rate levels and trends in organizational research. 2008;61:1139–60. doi: 10.1177/0018726708094863

42. Guy R, King E, Ayers A. The role of transvaginal ultrasound in the assessment of the female pelvis. Clin Radiol 1988;39:669–72. doi: 10.1016/S0009-9260(88)80090-4

43. Lefringhouse JR, Neward E, Ueland FR, Baldwin LA, Miller RW, DeSimone CP, et al. Probability of fallopian tube and ovarian detection with transvaginal ultrasonography in normal women. Womens Health (Lond) 2016;12:303–11. doi: 10.2217/WHE.15.111

44. Glanc P, Brofman N, Kornecki A, Abrams J, Farine D, Salem S. Visualization of the ovaries in early pregnancy by transvaginal sonography. J Obstet Gynaecol Can 2007;29:228–31. doi: 10.1016/S1701-2163(16)32415-X

45. Evans KD. Performance evaluation for the sonographer. 2016;9:143–6. doi: 10.1177/875647939300900307

46. Van Doorn LC, Dijkhuizen FP, Kruitwagen RF, Heintz AP, Kooi GS, Mol BW. Accuracy of transvaginal ultrasonography in diabetic or obese women with postmenopausal bleeding. Obstet Gynecol 2004;104:571–8. doi: 10.1097/01.AOG.0000136080.55874.7F

47. Malhotra N, Bahadur A, Singh N, Kalaivani M, Mittal S. Does obesity compromise ovarian reserve markers? A clinician's perspective. Arch Gynecol Obstet 2013;287:161–6. doi: 10.1007/S00404-012-2528-7

48. Halawaty S, ElKattan E, Azab H, ElGhamry N, Al-Inany H. Effect of obesity on parameters of ovarian reserve in premenopausal women. J Obstet Gynaecol Can 2010;32:687–90. doi: 10.1016/S1701-2163(16)34573-X

49. Melissant HC, Verdonck-de Leeuw IM, Lissenberg-Witte BI, Konings IR, Cuijpers P, Van Uden-Kraan CF. 'Oncokompas,' A web-based self-management application to support patient activation and optimal supportive care: a feasibility study among breast cancer survivors. Acta Oncol 2018;57:924–34. doi: 10.1080/0284186X.2018.1438654

50. Ethics Committee of the American Society for Reproductive Medicine. Disparities in access to effective treatment for infertility in the United States: an Ethics Committee opinion. Fertil Steril 2015;104:1104–10. doi: 10.1016/J.FERTNSTERT.2015.07.1139

51. Reindollar RH. Increasing access to infertility care-what will it take? Fertil Steril 2017;108:600–1. doi: 10.1016/J.FERTNSTERT.2017.09.002

52. The use of antimüllerian hormone in women not seeking fertility care. ACOG Committee Opinion No. 773. American College of Obstetricians and Gynecologists. Obstet Gynecol 2019;133:e274–8. doi: 10.1097/AOG.0000000000003162

Authors' Data Sharing Statement

Will individual participant data be available (including data dictionaries)? *Yes, but anonymized.*

What data in particular will be shared? *Analysis data sets used to generate tables of results in the manuscript.*

What other documents will be available? *None.*

When will data be available (start and end dates)? *From the date of publication up to 1 year.*

By what access criteria will data be shared (including with whom, for what types of analyses, and by what mechanism)? *Data and associated documentation will be available under the auspices of the Principal Investigator through a data-sharing agreement that provides for: i) IRB approval for the proposed research using the data; ii) a commitment to using the data only for research purposes and only for the IRB approved research, iii) a commitment not to try to identify any individual participant; iv) a commitment to securing the data using appropriate password protected computer technology and other necessary safeguards; v) a commitment to not transfer the data to other users and destroying the data after analyses are completed.*

PEER REVIEW HISTORY

Received September 25, 2021. Received in revised form November 23, 2021. Accepted December 3, 2021. Peer reviews and author correspondence are available at http://links.lww.com/AOG/C595.