Journal of Medical Systems (2022) 46: 38
https://doi.org/10.1007/s10916-022-01816-0

CLINICAL SYSTEMS

Within Clinic Reliability and Usability of a Voice-Based Amazon Alexa Administration of the Patient Health Questionnaire 9 (PHQ 9)

Jason Beaman[1] · Luke Lawson[1] · Ashley Keener[1] · Michael L. Mathews[2]

Received: 11 January 2022 / Accepted: 7 April 2022 / Published online: 10 May 2022

Abstract

Over the last two decades, metric-based instruments have garnered popularity in mental health. Self-administered surveys, such as the Patient Health Questionnaire 9 (PHQ 9), have been leveraged to inform treatment practice of Major Depressive Disorder (MDD). The aim of this study was to measure the reliability and usability of a novel voice-based delivery system of the PHQ 9 using Amazon Alexa within a patient population. Forty-one newly admitted patients to a behavioral medicine clinic completed the PHQ 9 at two separate time points (first appointment and one-month follow up). Patients were randomly assigned to a version (voice vs paper) completing the alternate format at the next appointment. Patients additionally completed a 26-item User Experience Questionnaire (UEQ) and open-ended questionnaire at each session. Assessments between PHQ 9 total scores for the Alexa and paper version showed a high degree of reliability ($\alpha = .86$). Quantitative UEQ results showed significantly higher overall positive attitudes towards the Alexa format with higher subscale scores on attractiveness, stimulation, and novelty. Further qualitative responses supported these findings with 85.7% of participants indicating a willingness to use the device at home. With the benefit of user instruction in a clinical environment, the novel Alexa delivery system was shown to be consistent with the paper version giving evidence of reliability between the two formats. User experience assessments further showed a preference for the novel version over the traditional format. It is our hope that future studies may examine the efficacy of the Alexa format in improving the at-home clinical treatment of depression.

Keywords Depressive disorder, major · Patient health questionnaire · Voice recognition · Amazon Alexa · Mental health · IoT

Introduction

Treatment background

Major Depressive Disorder (MDD) affects millions of individuals in the United States. With an annual incidence of 10.4% and a lifetime prevalence of 20.6% [1], this condition ranks among the most common mental disorders. In 2017, the National Institute of Mental Health estimated 17.3 million people in the United States suffered at least one major depression episode [2]. Since the beginning of the COVID-19 pandemic, the rate of US adults experiencing symptoms of depression may have tripled [3]. Recent surveys have shown these upward trends continuing [4].

While MDD is not curable, it is treatable. Individuals diagnosed with MDD are monitored on a regular basis for treatment response and overall status. Monitoring has occurred in a few different ways. Traditional methods require in-person, monthly visits to medical clinics to determine treatment efficacy and/or fill prescriptions. Visits may be with a mental health provider or primary care physician. During these visits, clinicians assess the patient's depression by evaluating symptom severity with questions mimicking DSM criteria. However, responses are not quantified using traditional methods. Instead, the clinician uses this approach to make an individualized assessment of the patient's status, i.e., conducts a risk/benefit evaluation in relation to treatment options. Over the last two decades, a second method has garnered popularity and relies on self-administered surveys that ask patients to rate individual symptoms.

The use of surveys to monitor progress has the added benefit of introducing metric-based instruments. These

This article is part of the Topical Collection on *Clinical Systems*

✉ Luke Lawson
luke.lawson@okstate.edu

[1] Center for Health Sciences, Oklahoma State University, Tulsa, USA

[2] Oral Roberts University, Tulsa, USA

instruments, such as the Hamilton Depression Screen or Patient Health Questionnaire 9 (PHQ 9), are standardized tools that track patient improvement. Since its publication in 2001, the PHQ 9 has been prolific in the screening and/or monitoring of major depression with over 11,000 citations [5, 6]. Responses to the questionnaire are graded, i.e., the total score can be used to indicate mild, moderate, or severe episodes as it relates to the patient's major depression. These grades have an 88% consistency rate with clinical diagnoses of major depression severity [7]. Further, it can be utilized to determine whether or not the patient has improved since the last visit to better inform the treatment plan.

Current American Psychological Association (APA) guidelines for the treatment of MDD encourage utilizing standardized instruments, such as the PHQ 9, for initial patient psychiatric assessment, determining and implementing therapeutic protocols, and monitoring treatment progress [8]. Further recommendations from the US Preventative Task Force (USPTF) include screening all adults over the age of 18 with a standardized instrument and providing further assessment for those who screen positive [9]. This approach, supported through the Texas Medication Algorithm Project, has shown improved patient outcomes after one year, such as better clinical adherence to prescribing medications [10]. Beyond treatment benefits, the PHQ 9 has been used to better understand potential differences in MDD among differing patient populations. This includes, but is not limited, to differences between clinic and non-clinic populations [11], predisposition of certain sexual and gender groups for meeting certain criteria for MDD [12], and differences in diagnostic accuracy in patients diagnosed with cancer. [13].

Technological background

The advent of modern technologies in the past 10 years, including voice-based assistance, machine learning, and improvements in health information systems, can allow clinicians to assess patients regularly in their own home. For instance, devices equipped with the Amazon Alexa software enable users to control smart home devices through voice commands and upload data to cloud-based databases. Since 2014, the software has become ubiquitous with applications that can be enabled with smart TVs and cellular devices.

The COVID-19 pandemic has highlighted the need for at-home clinical care. Further, individuals suffering from MDD experience symptoms (e.g., loss of energy) that make it difficult to leave their home; such factors may contribute to high rates of patient attrition [14–16]. Therefore, the capacity to treat patients from home has the potential to provide regular care and improve patient outcomes. The advancement of remote clinical care, however, complicates the collection of diagnostic, survey-based information that has become important for depression treatment. While the traditional paper version of the PHQ 9 is convenient to administer at monthly in-clinic appointments, it can be very difficult to collect from patients at-home. Despite the general acceptance of the PHQ 9 within psychiatric clinics, the construct validity and user experience of mechanisms designed to collect these responses from home are largely unstudied. To overcome this barrier, our team has developed an app, using Amazon Alexa, that collects auditory responses from patients and stores those responses on a cloud database.

Significance

Making an efficient delivery system for quantitative measures is a logical next step for at-home clinical treatment. Internet of Things (IoT) technology is a field connecting household devices to the internet and has the potential to change the landscape of healthcare treatment [17]. The introduction of IoT monitoring devices within treatment for cardiovascular diseases has been shown to reduce fatality rates and costs incurred by patients suffering from heart failure [18]. While numerous mental health IoT studies have focused on data analytics and machine learning techniques to make predictions about future outcomes [19], the aim of this research is to pragmatically leverage the cloud capabilities of Amazon Alexa devices and provide clinicians with measures they are familiar with interpreting. Studies have shown a reduction in readmission rates to psychiatric hospitals when patients struggling with Bipolar disorder utilize smartphone based self-assessment questionnaires [20]. Due to the personal nature of the questions within the PHQ 9 and the social stigma of being diagnosed with depression, uncontrolled social settings could elevate patient response bias. IoT devices that utilize cloud storage capabilities are a promising technological area mitigating such obstacles.

There are three benefits of choosing an Amazon Alexa device as the platform for the IoT app: 1) Alexa devices are utilized consistently in household environments, 2) Amazon offers numerous IoT solutions that could be used for future treatments, 3) and the novelty of the devices may encourage response rates. However, it is unclear whether patients will answer mental health, voice-based questionnaires that contain sensitive questions in the same manner as paper-based questionnaires. In other words, responding to items "outloud" may elicit negative emotions, such as insecurity, and introduce further response bias. A review by Sezgin et al. [21] investigated the readiness of voice assistance devices in a clinical setting and a further review by Kumah-Crystal [22] investigated the usability of these devices in a general setting. Some concerns raised by both studies were the usability of the devices and the potential of patients to adjust their "voice responses" to get through a mandated questionnaire.

Purpose of the study

The purpose of this study was two-fold: 1) examine the reliability of Amazon Alexa within a clinical environment utilizing a standardized instrument, i.e., PHQ 9, and 2) evaluate patients' attitudes towards the novel delivery system. Prior to testing the efficacy of this new diagnostic tool at-home, we aim to provide evidence that the tool is clinically reliable and patients are comfortable with the device. Positive patient reactions to a reliable Amazon Alexa format of the PHQ 9 has the potential to bridge the gap between in-clinic and at-home care for patients suffering from depression. It is our hope the findings from this study will lay the groundwork for future investigations and leverage the potential of interactive voice assessments to improve at-home clinical care.

Methods

Participants

Participants in the study were newly admitted, in-person patients (ages 18 years or older) at a behavioral medicine clinic. Vulnerable populations, such as children (i.e., minors or individuals under the legal age of consent) and individuals who are incarcerated (i.e., prisoners) were excluded. Additionally, individuals who are not their own guardian (i.e., those suffering from severe disabilities) were also excluded. A total of 81 patients were approached and asked to participate in the study; 56 (69%) chose to participate in one session and 41 (51%) completed both sessions. Demographic characteristics of the patients who completed both baseline and follow-up sessions can be found in Table 1 below.

Measures

Patient health questionnaire–9 The PHQ-9 is a 9-item depression module that measures major depression. Patients were expected to respond to Likert-type items ranging from 0 (*not at all*) to 3 (*nearly every day*). Example items from the scale include "little interest or pleasure in doing things" and "poor appetite or overeating." In terms of severity, scores can range from 0 to 27. An individual can be diagnosed as experiencing major depression if 5 or more of the 9 depressive symptom criteria have been present at least "more than half the days" in the past 2 weeks. Additionally, 1 of the symptoms must be depressed mood or anhedonia. If 2, 3, or 4 depressive symptoms have been present in the past 2 weeks "more than half the days," and 1 of the symptoms is depressed mood or anhedonia, then other forms of depression may be diagnosed [5]. Aggregate scores can further be interpreted from the surveys with scores 1–4 corresponding to minimal depression, 5–9 corresponding to mild depression, 10–14 corresponding to moderate depression, 15–19 corresponding to moderately severe depression, and 20–27 corresponding to severe depression [7]. The participants' responses to each of the 9 items were summed for each patient session to form a composite score.

User experience questionnaire The UEQ measures a user's experience in relation to a product. The questionnaire utilizes a semantic differential as its item format. In other words, each item consists of a pair of terms with opposite meanings that can be rated on a 7-point Likert scale. For instance, one example item includes the terms "not understandable" to "understandable" where a 1 would indicate the user fully agrees with the negative term and a 7 would

Table 1 Demographic Characteristics for patients who completed sessions by group assignment (Paper-Alexa, Alexa-Paper, and Total)

Parameter	Participants					
	Paper-Alexa		Alexa-Paper		Total	
	N	%	*N*	%	*N*	%
Female	15	71	10	50	25	61
Male	6	29	9	45	15	37
Transgender-Male	0	0	1	5	1	2
	N	%	*N*	%	*N*	%
White / Caucasian	15	17	13	65	28	68
Black / African American	1	5	1	5	2	5
Asian or Pacific Islander	0	0	0	0	0	0
Hispanic / Latino	0	0	0	0	0	0
American Indian	5	24	3	15	8	20
Multiple	0	0	2	10	2	5
Prefer Not to Answer	0	0	1	5	1	2
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age (yrs)	39.0	16.7	44.5	15.0	41.7	15.9

indicate the user fully agrees with the positive term. The UEQ contains a total of 26 items where 13 of the items are reverse coded. Further, a total score and 6 subscale scores (attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty) can be formed [23, 24]. The attractiveness subscale contains 6 items whereas the other subscales consist of 4 items each. The survey includes 26 items composed of items from each of the six subscales (attractiveness, efficiency, perspicuity, dependability, stimulation, and novelty) (see Table 2). Responses were reverse coded to account for valence of the question and summed for each patient to form a composite score for both the overall user experience and for individual subscales.

Supplemental questionnaire A qualitative questionnaire was provided to patients after the session where they completed the PHQ 9 in the Amazon Alexa format. The supplemental questionnaire included 2 additional questions: 1) Would you be willing to use the device at home and 2) Do you have any additional comments regarding your experience with the device? The first question was a forced choice where participants were asked to respond "yes," "unsure," or "no" while the second was open-ended.

Procedure

Newly admitted patients were asked to complete the questionnaires (i.e., PHQ-9 and the UEQ) at two different time points (i.e., Baseline and 1-month follow-up). In order to control for order effects, the procedure was counterbalanced. To achieve this goal, patients were randomly assigned to two different groups. Patients assigned to the first group (Paper-Voice) completed the PHQ-9 on the paper format at the baseline session and then completed the PHQ 9 in the Amazon Alexa format approximately at their usual follow-up appointment approximately one month later. Those assigned to the second group (Voice-Paper) completed the PHQ-9 on voice-based format at baseline and then completed the paper format at the 1-month follow-up. UEQ were completed at each time point with each patient having a separate UEQ for both the paper and voice-based format. The supplemental questionnaire was administered immediately after patients completed their Amazon Alexa PHQ 9 assessment.

Statistical analyses

Patient demographic backgrounds were assessed between groups through univariate ANOVAs on gender, age, and ethnicity, to determine potential imbalances in group assignment distribution. Possible interactions of PHQ 9 overall scores between group assignment (Paper-Voice vs Voice-Paper) and appointment sessions (baseline vs follow-up) were examined through a two-way ANOVA. After possible interactions were observed, a within patient Cronbach's alpha was calculated to assess the internal consistency between the Amazon Alexa and paper format. In line with other interpretations of the Cronbach's alpha coefficient [25, 26], a value between 0.75 and 0.95 was interpreted as having a high degree of internal consistency.

A 2×2 design was used for the analysis of the total user experience score after reverse coding for negatively valanced Likert questions. Group assignment (paper-voice vs voice-paper) was treated as the between-subjects factor and appointment session (baseline vs follow-up) was accounted for in a repeated measures ANOVA statistical design. Further analyses were conducted to test for differences between the six subscales. The six subscales of the UEQ (attraction, perspicuity, novelty, stimulation, dependability, and efficacy) were treated as the univariate dependent variables in a repeated measures MANOVA with the omnibus interaction testing whether responses to any of the subscales differed from one another. Statistical differences of the six subscales of the UEQ were examined as the dependent variables of a profile analysis using group assignment and session as the independent factors. Bartlett's test and Box's M test were calculated for both the UEQ total scores and UEQ subscales to confirm assumptions had been met for calculating the omnibus test statistics while Levene's test was assessed for follow up univariate analyses. Effect sizes, such as partial eta squared, and confidence intervals were additionally calculated. Significant interactions were followed up with simple effect analyses and statistically significant main effects were examined via Bonferroni multiple comparison procedures. All statistical analyses listed including measures for internal consistency of the PHQ 9, the repeated measures ANOVA UEQ total score and profile analysis MANOVA

Table 2 UEQ subscales: title, summary of corresponding item number and item questions

Subscale Title	Subscale Theme	Item Number
Attractiveness	Do users like or dislike the product?	1, 12, 14, 16, 24, 25
Efficiency	Is the product efficient and well organized?	9, 20, 22, 23
Perspicuity	Is the product intuitive and easy to learn?	2, 4, 13, 21
Dependability	Does the product seem secure and predictable?	7, 11, 17, 19
Stimulation	Is it exciting to use the product?	5, 6, 7, 18
Novelty	Is the design of the product innovative and creative?	3, 10, 15, 26

for UEQ subscales were conducted using SPSS version 27 (IBM SPSS 27).

Results

Demographics Analyses showed no differences in demographic variables between group assignments (paper-voice vs voice-paper) for gender (χ^2 (2, N = 41) = 2.58, $p = 0.28$), ethnicity ($\chi^2(5, N = 41) = 3.62$, $p = 0.61$), or age ($F(1,39) = 1.21$, $p = 0.28$). As no demographic differences were observed between group assignment, further analyses were conducted without adjustment.

Information was not collected on treatment plans within the clinic. PHQ 9 results do, however, provide further insight into patient background for the 41 patients who completed both sessions. In accordance with screening guidelines for the PHQ 9 [7], five patients were screened for minimal depression, eight for mild depression, 10 for moderate depression, 10 for moderately severe depression, and 8 for severe depression at their first session. At their second session, six patients were screened for minimal depression, 11 for mild depression, 13 for moderate depression, eight for moderately severe, and three for severe depression.

Patient health questionnaire–9 ANOVA analysis of PHQ 9 total scores revealed no significant interaction between group assignment and appointment session. Because no significant interactions were observed between appointment sessions, internal consistency between the paper and voice-based assessments was assessed without controlling for session effects and were measured through a Cronbach's alpha test. These analyses showed a high degree of reliability between paper and Amazon Alexa formats ($\alpha = 0.86$, 95% CI = 0.77, 0.94).

User experience questionnaire The interaction of assignment group and session of the 2 × 2 mixed model ANOVA was used to test whether patients had differing total user experience scores between the Amazon Alexa and paper formats. Assumptions for the repeated measures ANOVA were assessed through Bartlett's test and Box's M test. Bartlett's test was significant (approximate $\chi^2 = 192$, df = 15, $p < 0.00$) and Box's M failed to reach significance, both conditions meeting the overall assumptions for the omnibus MANOVA and justifying the use of a Wilk's Lambda test statistic. The omnibus MANOVA analysis of UEQ total scores revealed a significant interaction between group assignment and appointment session (Wilk's Lambda = 0.645, $F(1, 39) = 21.46$, $p < 0.00$, $\eta^2 = 0.175$). As shown graphically in Fig. 1, this finding indicates a higher overall user experience rating for both sessions where patients completed the PHQ 9 in the Alexa format (higher values for voice-paper group baseline session and paper-voice group follow-up session). No significant main effect for group assignment was observed between sessions indicating that there were no assignment or order effects for patients completing the questionnaire.

UEQ MANOVA (6 subscales) After revealing a significant interaction for UEQ total score, a repeated measures MANOVA was conducted to investigate any potential differences between subscales. Assumptions for the repeated measures MANOVA comparing the six subscales were assessed through Bartlett's test and Box's M test. Bartlett's test was significant (approximate $\chi^2 = 230.59$, df = 66, $p < 0.00$) consistent with the assumptions of a MANOVA. Box's M test was, however, significant as (Box's M = 154.44, $F(78, 4777) = 1.40$, $p = 0.01$) indicating that a Pillai's Trace test statistic may be most appropriate. The omnibus MANOVA analysis of UEQ subscales indicated that there is a significant interaction between session and group assignment on the six UEQ subscales (Pillia's Trace = 0.51, $F(5, 35) = 0.51$, $p < 0.01$) indicating that participants responded differently to certain subscales based on which format they interacted with.

With a significant omnibus MANOVA, follow up univariate tests were conducted to assess potential differences in each of the subscales. Levine's tests were conducted for each of the six subscales. None of the individual subscales reached a significant value for Levene's test indicating that each maintained an acceptable level of homogeneity of variance. After utilizing a Bonferroni adjustment, three of the subscales showed a significant interaction between group assignment and session indicating higher subscale scores for the Amazon Alexa format. These three subscales were those measuring attractiveness ($F(1,39) = 22.87$, $p < 0.00$, $\eta^2 = 0.37$), stimulation ($F(1,39) = 14.72$, $p < 0.00$, $\eta^2 = 0.27$), and novelty ($F(1,39) = 25.23$, $p < 0.00$, $\eta^2 = 0.39$). Each of these three interactions can be observed in Fig. 2.

Patient qualitative responses Of the 56 patients who completed at least one session, 49 patients (paper-voice = 21 and voice-paper = 28) completed a PHQ 9 on the Amazon Alexa format. Of these, 42 patients (85.7%) reported that they would be willing to use the device at home, three patients (6.1%) responded as unsure, and four patients (8.2%) responded that they would not be willing to use the device at home. Twenty-three patients provided further comments on their disposition towards the device. Of those who provided further comments, 19 patients responded as being willing to use the device at home, four responded as unwilling, and no one who responded as unsure provided a comment. Those who provided a comment represented 45.2% of those willing to use the device at home and everyone who was unwilling provided further comments.

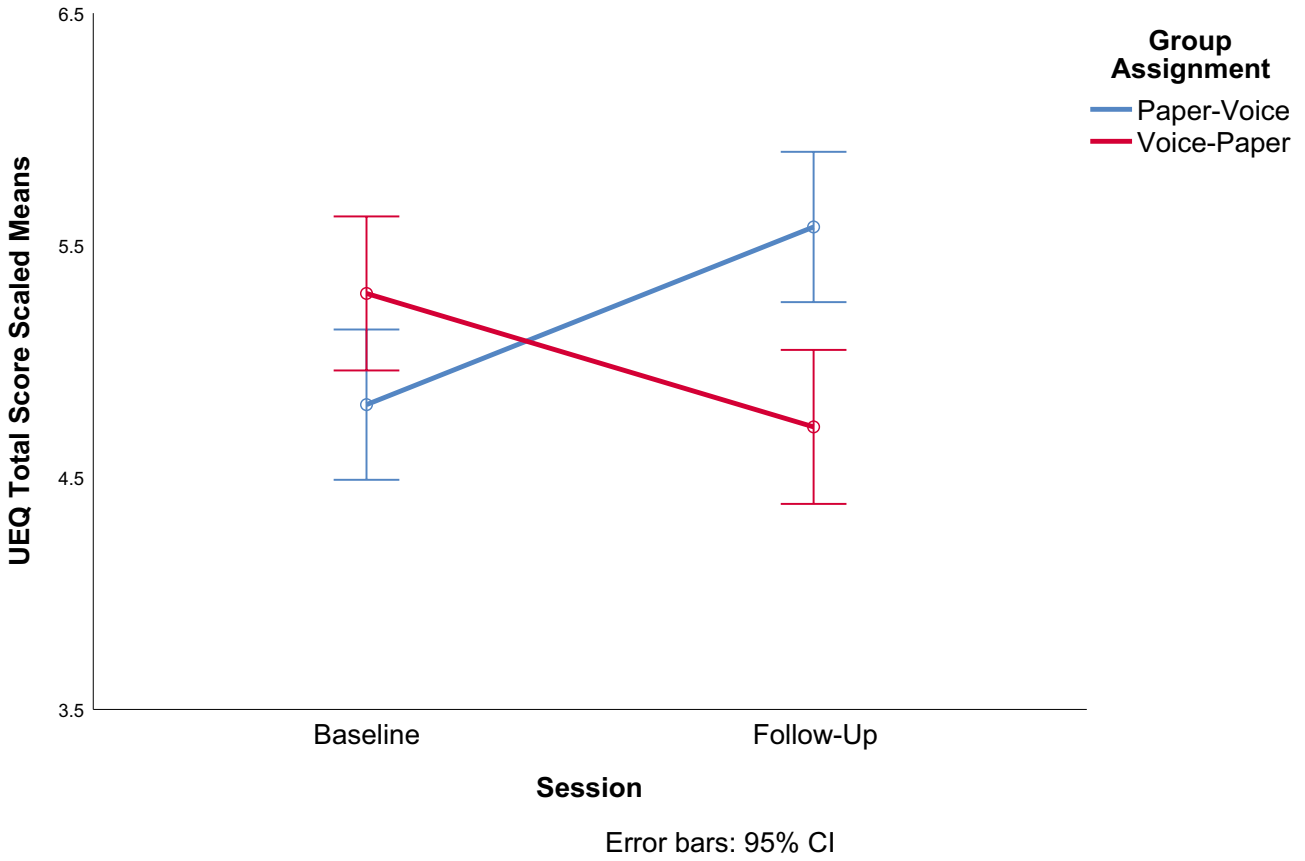

Fig. 1 Means for UEQ Total Score scaled for individual responses by Group Assignment (Paper-Voice and Voice-Paper) and Session (Baseline and Follow-Up)

Of the 23 comments, 13 (56.5%) provided feedback that expressed their excitement for the new technology and study such as "very pleasant and would recommend it to others" or "very much enjoyed the experience". Nine participants included feedback on the usability of the device with four identifying the ease of use and five identifying specific pain points in using the device such as "struggled understanding my voice" or "I like the mirror format but it takes WAY longer than just filling out a form". Of the four who expressed reluctance to use the device at home, three (75%) were concerned with the security of the Alexa device with the most colorful response being "I wouldn't have 'the mirror' or Alexis etc. at home regardless of this particular application. (Down with the Matrix!)".

A break-down of willingness to use the device at home by demographic backgrounds was inconclusive. Of the seven patients (14.3%) who were unwilling or unsure, five were female and two were male. One out of nine (11.1%) patients who identified as American Indian, two out of four (50%) who identified as Black or African American, and three out of 31 (9.7%) who identified as White / Caucasian were either unwilling or unsure whether they would use the device at home. The higher percentage of Black or African American responding negatively toward the device may reflect skepticism towards research observed in wider society [27], but the sample size of this study is far too small to make a conclusive statement.

Discussion

Our research team set out to test the utility of a novel diagnostic tool with potential to improve clinical treatment for patients suffering from depression. Before implementing the tool in a wider context, the reliability of the device in measuring depression and overall patient opinion needed to be assessed. If evidence were lacking for either of those questions, further research on the device would be unproductive. Findings from our study provide support that the administration of the PHQ 9 through a voice-based Alexa device was consistent with the paper version and was rated significantly higher regarding user experience. Specifically, patients rated the Alexa format significantly higher for attractiveness, level of stimulation, and novelty. While qualitative responses revealed some hesitance regarding security, no significant differences were observed between format type

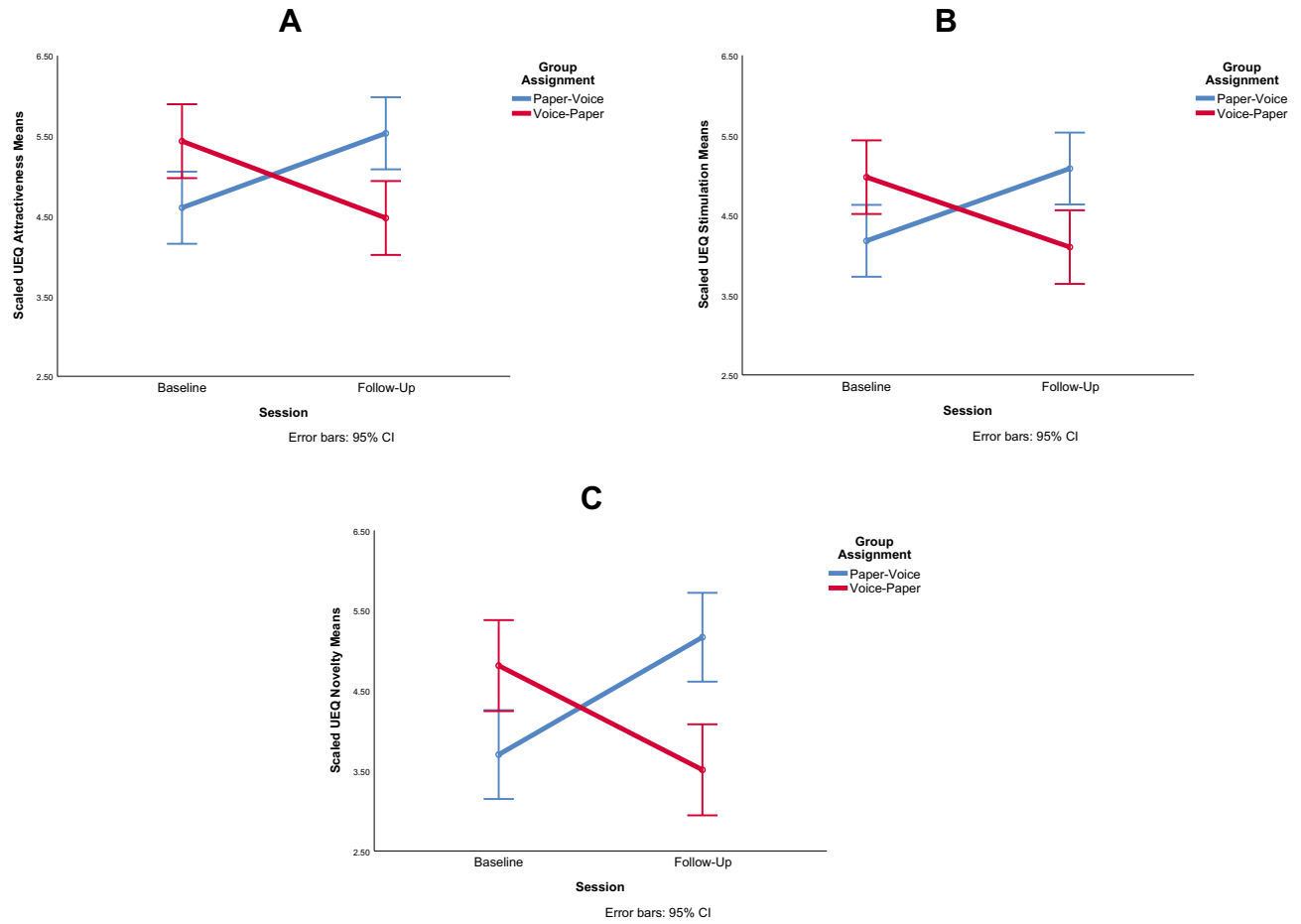

Fig. 2 Means of UEQ Subscales for Attractiveness (**A**), Novelty (**B**), and Stimulation (**C**) by group assignment and session

on the dependability of the device. Further, 85.7% of the patients recruited for the study stated they were willing to use the device at home.

Through demonstrating the consistency of the Amazon Alexa version and paper format, the study has provided evidence of the convergent validity of the voice-based PHQ 9. Stated in other words, the consistency of the Alexa and paper versions of the PHQ 9 provide early evidence that the voice-based version is reliably measuring depression. This consistency was demonstrated in one of the more critical cases of use for the PHQ 9 in the screening of new mental health patients. The positive reception of the device is further encouragement that the new delivery method of the assessment may be accepted outside of the study.

Limitations and future directions

Patient depression levels are not expected to be a static across time and should ideally improve as patients seek resources to treat their depression. One limitation of the study was that the reliability of the formats was compared across a one-month delay. This delay could introduce instability into the measurement of consistency between formats especially if patients being treated for depression see disproportionate gains in their depression level. Researchers chose to move forward with this research design compared to alternatives such as testing patients on both formats within the same session as they felt patients may bias their responses by attempting to replicate their answers from earlier in the appointment. They also felt testing within the same session may introduce additional patient user fatigue and irritability. A cross-balanced research design randomly assigning patients to a format at their first session and the alternative format at their second was chosen to attempt to mitigate these concerns.

An additional limitation of this study was the controlled nature in which patients interacted with the device. Patients were instructed on how to use the device and completed practice questions prior to using Amazon Alexa. Further, the device was already installed in the testing location. Ergo, such factors may alter a patient's user experience. Future studies should investigate the user's experience regarding device installation and within different environments. Additionally, longitudinal studies, comparing different delivery systems (e.g., video conferencing

therapy versus voice-based PHQ 9 assessments), in relation to treatment outcomes and attrition rates. Another direction could be to integrate the Alexa PHQ 9 assessment with additional IoT devices to provide more active treatment for patients both in clinic and at household settings. Examples of IoT capabilities that are already possible include the measurement of patients' posture [28] and breathing rates. With clinician guidance, tools could be designed to help patients treat their depression symptoms through active posture and breathing exercises.

Conclusion

The rates of patients suffering from depression in the U.S. have increased dramatically over the past five years [1–4]. While behavioral medicine clinics have relied on quantitative self-assessments for depression, such as the PHQ 9 [5, 6, 8, 9], the introduction of video conferencing technology has led to a gap in treatment with clinicians struggling to collect these metrics. Cloud based IoT solutions have shown numerous applications across healthcare [17, 19] and may bridge the gap between in-person and remote treatment. Findings from our study demonstrate 'good' internal consistency between a voice-based Amazon Alexa and paper-based version of the PHQ 9 within a clinical sample. The Amazon Alexa device was rated highly as attractive, stimulating, and novel. Further, a high proportion of patients indicated they would be willing to use the device at home. The logical next step is to test the device in an at-home environment. We hope that future studies will investigate the efficacy of the device outside the clinic and that more intricate treatment tools can be developed to serve in conjunction with the PHQ 9 app.

Declarations

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Conflict of interest Dr. Jason Beaman declares he has no conflict of interest. Luke Lawson declares he has no conflict of interest. Dr. Ashley Keener declares she has no conflict of interest. Michael Mathews declares he has no conflict of interest.

References

1. Hasin D., Sarvet A., Meyers J., Saha T., Ruan W., Stohl M., & Grant B. (2018) Epidemiology of Adult *DSM-5* Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry.* 75(4):336–346. https://doi.org/10.1001/jamapsychiatry.2017.4602
2. National Institute of Mental Health | United States (2017) Statistics for Major Depression 2017. Accessed 11 Sept 2020. https://www.nimh.nih.gov/health/statistics/major-depression.shtml
3. Ettman, C., Abdalla, S., Cohen, G., Sampson, L., Vivier, P. & Galea, S. (2020) Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic, *JAMA Network Open. 3(9):e2019686.* https://doi.org/10.1001/jamanetworkopen.2020.19686
4. Abbot, A. (2021, February 3) COVID's mental-health toll: how scientists are tracking a surge in depression. *Nature.* https://doi.org/10.1038/d41586-021-00175-z
5. Kroenke, K., Spitzer, R.L. & Williams, J.B.W. (2001) The PHQ-9. *Journal of General Internal Medicine, 16: 606-613.* https://doi.org/10.1046/j.1525-1497.2001.016009606.x
6. Kroenke, K. (2021). PHQ-9: global uptake of a depression scale. *World Psychiatry*, *20*(1), 135–136. https://doi.org/10.1002/wps.20821
7. Manea L., Gilbody S. & McMillan D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ JAMC. 184(3) E191-E196.* https://doi.org/10.1503/cmaj.110829
8. Alan J Gelenberg, Marlene P Freeman, John C Markowitz, Jerrold F Rosenbaum, Michael E Thase, Madhukar H Trivedi, Richard S Van Rhoads, Victor I Reus, J Raymond DePaulo Jr, Jan A Fawcett, Christopher D Schneck, & David A Silbersweig. (2010). Practice Guideline for the Treatment of Patients With Major Depressive Disorder, Third Edition. The American Journal of Psychiatry, 167(10), 1–. Accessed 28 Oct 2021. https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf
9. Siu A. and the US Preventive Services Task Force (USPSTF) (2016) Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA.* 2016;315(4):380–387. https://doi.org/10.1001/jama.2015.18392
10. Crismon, M. L., Trivedi, M., Pigott, T. A., Rush, A. J., Hirschfeld, R. M. A., Kahn, D. A., DeBattista, C., Nelson, J. C., Nierenberg, A. A., Sackeim, H. A., & Thase, M. E. (1999). The Texas medication algorithm project: Report of the Texas consensus conference panel on medication treatment of major depressive disorder. *The Journal of Clinical Psychiatry*, *60*(3), 142–156. https://doi.org/10.4088/JCP.v60n0302
11. Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., & Horikoshi, M. (2018). Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. *PloS One, 13(7), e0199235–e0199235*. https://doi.org/10.1371/journal.pone.0199235
12. Borgogna, N. C., Brenner, R. E., & McDermott, R. C. (2021). Sexuality and gender invariance of the PHQ-9 and GAD-7: Implications for 16 identity groups. *Journal of Affective Disorders*, *278*, 122–130. https://doi.org/10.1016/j.jad.2020.09.069
13. Hartung, T. J., Friedrich, M., Johansen, C., Wittchen, H., Faller, H., Koch, U., Brähler, E., Härter, M., Keller, M., Schulz, H., Wegscheider, K., Weis, J., & Mehnert, A. (2017). The Hospital Anxiety and Depression Scale (HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as screening instruments for depression in patients with cancer. *Cancer*, *123*(21), 4236–4243. https://doi.org/10.1002/cncr.30846
14. Fava, M., & Kendler, K. S. (2000). Major depressive disorder. *Neuron (Cambridge, Mass.)*, *28*(2), 335–341. https://doi.org/10.1016/S0896-6273(00)00112-4
15. Warden, D., Rush, A.J., Carmody, T.J., Kashner, T.M., Biggs, M., Crismon, M.L. & Trivedi, M. (2009) Predictors of Attrition During One Year of Depression Treatment: A Roadmap to Personalized Intervention. *Journal of Psychiatric Practice.* https://doi.org/10.1097/01.pra.0000348364.88676.83

16. Mohr, D., Vella, L., Hart, S., Heckman, T., & Simon H (2008). The Effect of Telephone-Administered Psychotherapy on Symptoms of Depression and Attrition: A Meta-Analysis. *Clinical Psychology: Science and Practice*. https://doi.org/10.1111/j.1468-2850.2008.00134.x
17. Habibzadeh, H., Dinesh, K., Shishvan, O. R., Boggio-Dandry, A., Sharma, G., & Soyata, T. (2019). A survey of healthcare Internet of Things (HIoT): A clinical perspective. *IEEE Internet of Things Journal*, 7(1), 53-71. https://doi.org/10.1109/JIOT.2019.2946359
18. Schmier, J. K., Ong, K. L., & Fonarow, G. C. (2017). Cost-Effectiveness of Remote Cardiac Monitoring With the CardioMEMS Heart Failure System. *Clinical Cardiology (Mahwah, N.J.)*, *40*(7), 430–436. https://doi.org/10.1002/clc.22696
19. de la Torre Díez, I., Alonso, S. G., Hamrioui, S., Cruz, E. M., Nozaleda, L. M., & Franco, M. A. (2019). IoT-Based Services and Applications for Mental Health in the Literature. *Journal of Medical Systems*, *43*(1), 1–6. https://doi.org/10.1007/s10916-018-1130-3
20. Faurholt-Jepsen, M., Tønning, M., L., Frost, M., Martiny, K., Tuxen, N., Rosenberg, N, Busk, J., Winther, O., Thaysen-Petersen, D., Aamund, K., A., Tolderlund, L., Bardram, J., E., Vedel, L. (2020) Kessing Reducing the rate of psychiatric Re-ADMISsions in Bipolar Disorder using smartphones The RADMIS trial. *Acta Psychiatrica Scandinavica*. https://doi.org/10.1111/acps.13274
21. Sezgin, E., Huang, Y., & Ramtekkar, U. (2020) Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *NPJ Digit. Med.* 3, 122. https://doi.org/10.1038/s41746-020-00332-0
22. Kumah-Crystal Y., Pirtle C., Whyte H., Goode E., Anders S., & Lehmann C. (2018) Electronic Health Record Interactions through Voice: A Review. *Appl Clin Inform*. July; 9(Abbot, 2021):541–552. https://doi.org/10.1055/s-0038-1666844
23. Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. *In HCI and Usability for Education and Work (pp. 63–76). Springer Berlin Heidelberg*. https://doi.org/10.1007/978-3-540-89350-9_6
24. Rauschenberger, M., Schrepp, M., Cota, M. P., Olschner, S., & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (UEQ). *International Journal of Artificial Intelligence and Interactive Multimedia*. https://doi.org/10.9781/ijimai.2013.215
25. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.*Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555
26. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
27. Scharff, D. P., Mathews, K. J., Jackson, P., Hoffsuemmer, J., Martin, E., & Edwards, D. (2010). More than Tuskegee: understanding mistrust about research participation. *Journal of health care for the poor and underserved, 21(3), 879–897*. https://doi.org/10.1353/hpu.0.0323
28. Lina Yao, Quan Z. Sheng, Wenjie Ruan, Tao Gu, Xue Li, Nick Falkner, & Zhi Yang. (2015). RF-Care: Device-Free Posture Recognition for Elderly People Using A Passive RFID Tag Array. *EAI Endorsed Transactions on Ambient Systems, 2(6),* 120. https://doi.org/10.4108/eai.22-7-2015.2260064