Expanding Access to Depression Treatment in Kenya Through Automated Psychological Support: Stage 2 Registered Report

Eric P. Green[1], Nicholas Pearson[1,2], Sathy Rajasekharan[2], Michiel Rauws[3], Angie Joerin[3], Edith Kwobah[4], Christine Musyimi[5], Rachel Jones[2], Chaya Bhat[1], Yihuan Lai[1], Antonia Mulinge[2], & Eve S. Puffer[1,6]

[1] Duke Global Health Institute

[2] Jacaranda Health

[3] X2AI

[4] Moi Teaching and Referral Hospital

[5] Africa Mental Health Foundation

[6] Department of Psychology and Neuroscience, Duke University

Author Note

MR is the CEO and Founder of X2AI and created Tess. AJ is an employee of X2AI. EPG is an unpaid advisor to the X2AI Ethical Advisory Board and has no financial stake in the company.

Correspondence concerning this article should be addressed to Eric P. Green, Duke Global Health Institute, Box 90519, Durham, NC, USA. E-mail: eric.green@duke.edu

Abstract

Background. Depression during pregnancy and in the postpartum period is associated with a number of poor outcomes for women and their children. Although effective interventions exist for common mental disorders that occur during pregnancy and the postpartum period, most cases in low- and middle-income countries go untreated because of a lack of trained professionals. Task-sharing models such as the *Thinking Healthy* Program have shown great potential in feasibility and efficacy trials as a strategy for expanding access to treatment in low-resource settings, but there are significant barriers to scale-up. We are addressing this gap by adapting *Thinking Healthy* for automated delivery via a mobile phone. This new intervention, *Healthy Moms*, uses an existing artificial intelligence system called Tess (Zuri in Kenya) to drive conversations with users.

Objective. The primary objective of this pre-pilot study was to gather preliminary data on the *Healthy Moms* perinatal depression intervention to learn how to build and test a more robust service. We did this through a single-case experimental design with pregnant women and new mothers recruited from public hospitals outside of Nairobi, Kenya.

Methods. We invited women to complete a brief, automated screening delivered via text messages to determine their eligibility. Enrolled participants were randomized to a 1- or 2-week baseline period and then invited to begin using Zuri. We prompted participants to rate their mood via short message service every 3 days during the baseline and intervention periods, and we used this preliminary repeated measures data to fit a linear mixed-effects model of response to treatment. We also reviewed system logs and conducted in-depth interviews with participants to study engagement with the intervention, feasibility, and acceptability. IRRID: DERR1-10.2196/11800.

Results. We invited 647 women to learn more about Zuri. Of those invited, 86 completed our automated SMS screening, and 41 enrolled in the study. Most of the enrolled women submitted at least 3 mood ratings (75.6%) and sent at least 1 message to Zuri (65.9%). A third of the sample engaged beyond registration (34.1%). The average woman who engaged with Zuri post-registration started and completed 3.4 (SD=3.2) and 3.1 (SD=2.9) *Healthy Moms* sessions, respectively. Most interviewees who had tried Zuri had a very positive attitude towards the service and expressed that they could trust Zuri. They also attributed positive life changes to the intervention. We estimated that using this alpha version of Zuri led to a 7% improvement in mood.

Conclusions. Zuri is feasible to deliver via SMS and was acceptable to this sample of pregnant women and new mothers. The results of this pre-pilot will serve as a baseline for future studies in terms of recruitment, data collection, and outcomes. The next step in Zuri's development is to refine the intervention content and add Swahili language support. Conversational agents like Zuri have great potential to address the large treatment gap that exists in many low-resource settings, both as a new channel of treatment and as an adjunct to traditional and task-shifting approaches.

Keywords: telemedicine; mental health; depression; artificial intelligence; Kenya; text messaging; chatbot; conversational agent

Expanding Access to Depression Treatment in Kenya Through Automated Psychological Support: Stage 2 Registered Report

Introduction

Depression is a leading cause of disability worldwide. Women suffering from perinatal depression are a particularly underserved population. Depression during pregnancy and in the postpartum period (perinatal depression) affects as many as 20% of women in high-income countries [1] and maybe more in low-income countries [2]. The condition is associated with a number of poor outcomes for women and their children, including increased maternal morbidity and mortality [3,4], poor infant health [5–9], and poor developmental outcomes [10–12].

Although effective interventions exist for common mental disorders that occur during pregnancy and the postpartum period [13], most cases in low- and middle-income countries (LMICs) go untreated. In these settings, more than 7 out of 10 people who need treatment cannot access care because of a lack of trained professionals [14]. In Kenya, for example, there are only 180 psychiatric nurses outside of the capital city, a ratio of 1 provider per 200,000 people. To close this gap, the World Health Organization developed the Mental Health Gap Action Programme intervention guide outlining how to deliver mental health services in primary health care settings through nonspecialist providers. This task-sharing approach has proven efficacious, particularly for maternal mental health [15].

A prime example is the 15-session *Thinking Healthy Program*, a cognitive behavior therapy (CBT)–based intervention for treating perinatal depression that is intentionally nonstigmatizing [16]. Community health workers—typically women educated through secondary school with no specific background in mental health—are trained over 5 to 10 days to help pregnant women learn three skills: (1) to identify unhealthy thinking, (2) to replace unhealthy thinking with helpful thinking, and (3) to practice thinking and acting healthy. In a trial in Pakistan with 900 pregnant women, Rahman et al found that the intervention halved the prevalence of major depression [17], and a 7-year follow-up study reported some spontaneous recovery among the control group but also a persistent effect of treatment [18]. A peer-delivered version of *Thinking Healthy* offers a potential cost-effective first-line strategy for treating perinatal depression [19]

Despite this impressive evidence of feasibility and efficacy, however, there are significant barriers to scale-up [20], and there is evidence that intervention effects might not extend to children of depressed mothers without additional engagement [21]. Common implementation challenges of task-sharing models such as *Thinking Healthy* include a lack of funding and infrastructure for training and service delivery, workforce retention in the absence of compensation or incentives for nonspecialists, high workloads, transportation costs, appointment scheduling logistics, and inadequate clinical supervision [22]. Although it is critical to study how to optimize and scale these task-sharing approaches, the fact remains that, today, most women in LMICs who need treatment still have no access to care.

Given this demand and barriers to scale-up, our intention is to make it possible for anyone with a basic phone to receive high-quality, evidence-based psychological support anytime, anywhere. We are attempting this in the context of perinatal depression by adapting *Thinking Healthy* to an existing artificial intelligence (AI) system for automated psychological support called Tess (which we have named *Zuri* in Kenya). This idea is innovative because it introduces an entirely new delivery channel that has the potential for a step change in expanding access to care, while also potentially augmenting and strengthening existing task-sharing models.

Zuri works by engaging a patient in conversation via a variety of trusted channels, including text messaging (short message service, or SMS). Either Zuri or the patient can start a conversation, and Zuri can be programmed to walk a patient through a structured curriculum such as *Thinking Healthy.* As a safety measure, conversations with patients in need of additional support can be handed over to live counselors as needed. Potential benefits of this approach include on-demand 24/7 access for an unlimited number of patients, no scheduling of appointments, no travel costs to appointments, enhanced sense of privacy and avoidance of social stigma, and high fidelity to treatment.

Study Objectives

Our long-term goal is to expand access to high-quality, on-demand treatment services to people who suffer from common mental disorders such as perinatal depression but cannot receive care from mental health professionals because of cost and human resource constraints. The main objectives of this study were to adapt *Thinking Healthy* for dissemination in Kenya through the Zuri AI system; develop and test study procedures to inform the design of a randomized controlled trial; and generate preliminary evidence of feasibility, acceptability, and response to treatment.

Methods

Research Design

We adapted *Thinking Healthy* for the Zuri AI system and evaluated the combined perinatal depression intervention (which we are calling *Healthy Moms*) with a cohort of pregnant women and new mothers recruited from two large public hospitals in Kenya. We used a single-case experimental design (partially nonconcurrent multiple baseline [23], open label) and qualitative interviews to generate preliminary data on feasibility, acceptability, and response to treatment. This is a Stage 2 Registered Report. The Stage 1 protocol (DERR1-10.2196/11800) describes our preliminary work to adapt *Thinking Healthy* for dissemination in Kenya through the Zuri AI system [24].

Participants and Recruitment

We recruited pregnant women and new mothers from two large public hospitals in Kiambu County, Kenya. Both hospitals are part of a county-wide partnership offering patients innovative SMS programs that promote healthy motherhood [25]. When a woman signed up for the county SMS service, we sent her an invitation via SMS to complete an automated SMS screening (in English) to determine if she was eligible for *Healthy Moms*. The screening included questions about age, maternity status, expected or actual delivery date, 9 questions about symptoms of depression from the Patient Health Questionnaire-9 (PHQ-9) [26], and a question about her current mood.

We informed all women who completed the automated screening that a study team member would call them within 1 business day. During this follow-up call, women who endorsed having thoughts of self-harm in the previous 2 weeks (Question 9 on the PHQ-9) were offered a referral for counseling but were not eligible to enroll in *Healthy Moms* given the early stage of intervention development. All other women were eligible to enroll as long as they confirmed that they were at least 20 weeks pregnant or no more than 6 months postpartum. The study coordinator (AM)—a Kenyan woman fluent in English and Swahili—assessed each woman's English-speaking ability on the call and asked women to rate their ability to read and understand English. Women could enroll regardless of language ability, but we informed low English literacy women that they might not find value in the current version of the program if they were not comfortable reading and writing in English.

If a woman chose to continue the enrollment process, the study coordinator read the informed consent form, answered her questions, and obtained verbal informed consent to enroll. She asked enrollees to share information about the type of phone they use, schooling, number of dependents, marital status, and employment status.

Eligibility

To be eligible to participate, women needed to meet the following criteria: (1) pregnant (>20 weeks) or less than 6 months postpartum; (2) receiving antenatal or postnatal health care services from a participating hospital in Kiambu County; (3) enrolled in the county SMS program; and (4) at least 18 years of age. English language proficiency and self-reported experience of depression symptoms were not required but were assessed. Women who endorsed suicidal ideation at the time of recruitment were ineligible to enroll in the study and were informed about potential resources for treatment.

Randomization to Baseline Length

As each woman enrolled in the study, we attempted to match her to another new enrollee of similar maternity status and randomly assigned the pair to have a 1-week or 2-week baseline period (using a random number generator). The intention was to ensure that every participant had a concurrent baseline period with at least one other person.

Intervention

We invited women to participate in phone sessions of the *Healthy Moms* intervention based on their maternity status at enrollment. We modeled these automated SMS sessions after the original *Thinking Healthy* manual that was developed to guide community health workers to deliver the intervention in-person over 15 sessions [16]. We also created a companion *Healthy Moms* journal that we printed and delivered to enrolled participants [27]. The journal included modified health calendars from the original *Thinking Healthy* intervention along with short session summaries and writing prompts. This pilot study was an opportunity to get feedback on the journal to inform how we might adapt the content into text, audio, and video for electronic delivery (and ultimately discontinue print versions in future trials). We conducted an initial round of user testing to develop the SMS intervention journal content [28].

During each *Healthy Moms* session, women interacted with the automated system via SMS. Late in the study we also enabled women to chat with Zuri via Facebook Messenger. In between sessions, women were encouraged to start a conversation with Zuri by sending a free message. Zuri attempted to discern the user's request and responded automatically with answers or replies that employed active listening techniques such as restatement and reflection. If a woman discussed self-harm or other crisis topics, Zuri had the ability to alert a live study support member who could take over the chat session or call the participant directly and facilitate a referral to traditional in-person treatment if indicated (Zuri was programmed to inform women that her response might not be immediate at this stage of testing, so they should seek help at an emergency room if in a crisis). During enrollment we also informed participants that they were free to seek concomitant care and interventions at any point during the study.

Just as mental health specialists and nonspecialists trained to deliver psychotherapy improve over time with practice and experience, AI-enhanced systems such as Zuri also change, albeit in more subtle ways, given the current state of the technology. For instance, Zuri's emotion recognition algorithms updated automatically when it correctly or incorrectly interpreted the emotional valence of a user's input, but the didactic intervention content did not change dynamically. Modifications to the intervention content were made manually; we reviewed conversation transcripts and made minor changes to the wording or sequence of messages when we noticed that users were confused or not engaging.

Outcomes and Data Collection Procedures

We collected data on study implementation, intervention engagement, feasibility and acceptability of the intervention, and patient outcomes, including depression severity and current mood.

Study Implementation. We tracked data on the recruitment funnel from the initial screening invite through the secondary eligibility screening to ultimate engagement with the intervention. We also tracked participants' responses to regular prompts to complete automated assessments throughout the study period.

Intervention Engagement. We assessed intervention engagement by reviewing Zuri system logs to document completion of *Healthy Moms* sessions and patient-initiated engagement with Zuri outside of scheduled sessions. The Zuri system logs also informed our assessment of feasibility and acceptability; low engagement was considered a marker of potential barriers to feasibility or a lack of acceptability.

Feasibility and Acceptability of the Intervention. We further explored feasibility and acceptability by inviting 15 enrolled women to participate in individual interviews during the evaluation period. We purposively invited 3 different types of participants: those who did not finish the registration process with Zuri (5), those who finished the registration process but did not complete a session (5), and those who completed at least one session (5). A Master's-level trainee (YL) and the study coordinator (AM, Kenyan) conducted the interviews. Women who did not complete a full session with Zuri were interviewed by telephone. Women who completed one or more sessions were reimbursed to travel to one of the study hospitals for an in-person interview. The interviews lasted approximately 20 to 40 minutes and followed a semi-structured interview guide. The guide included open-ended questions and follow-up probes related to reasons for using Zuri, attitudes towards Zuri, favorite features, preferences of language and platform, challenges encountered, and perceived impacts after using Zuri. The interviews were conducted in English, but the study coordinator provided simultaneous translation to Swahili as needed.

In addition to these interviews, we also attempted to document all contact the research team had with participants outside of the Zuri AI system and logged all adverse events. We were interested in determining how much assistance or encouragement users need from the team to understand and use the automated intervention.

Patient Outcomes. To measure mood, we asked participants to rate their feelings on a 10-point scale that we created and tested with users [24], where 1 meant very sad and 10 meant very happy (shifted to 0-9 for analysis). We invited women to rate their current mood via SMS during the enrollment screening and then every 3 days throughout the baseline and intervention periods. Each rating invitation reminded women of their previous rating. We also encouraged women to track and reflect on their mood and behaviors on a daily basis using the *Healthy Moms* journal we provided as part of the intervention (not analyzed) [27].

We also administered the PHQ-9 [26] via SMS. Our intention was to assess depression severity throughout the intervention period, but after developing the protocol we determined that the depression screening was too long to administer on a repeating basis. Instead we opted to collect our minimum target of two self-ratings of depression severity representing pre- and post-treatment.

Empirical Approach

Describe Study Implementation and Intervention Engagement. We used the study database to summarize the recruitment funnel and outcome data collection progress. We quantified intervention engagement in several ways. First, we used the system logs to summarize how frequently each participant engaged with the intervention by either participating in a *Healthy Moms* session (in response to a scheduled invite) or initiating a chat with Zuri in between scheduled sessions. We also calculated and summarized the delay between our invitations to begin a *Healthy Moms* session and participants' start times, the proportion of *Healthy Moms* sessions started and completed, and the duration of participant-initiated chats with Zuri.

Explore Intervention Feasibility and Acceptability. As a hypothesis-generating exercise, we estimated the magnitude and direction of associations between participant characteristics measured at baseline (e.g., age, education, literacy, and symptom severity) and intervention engagement by fitting a Bayesian linear regression model.

We also explored barriers to and facilitators of engagement during in-depth interviews with participants and reviews of chat transcripts. Throughout the process, the interviewer/analyst (YL) wrote memos to capture the main themes. In preparation for the thematic analysis, she developed a codebook and randomly selected one transcript that was double-coded and discussed. After refining the codebook, the analyst used NVivo 12 to code memos and transcripts. The analyst wrote analytic memos for each thematic code, identifying similarities and differences across transcripts using a constant comparative method [29]. She identified representative quotations of each theme.

Generate Preliminary Evidence About Participant Response to Treatment. We aggregated the individual N-of-1 studies and estimated the magnitude of response and quantified uncertainty by fitting Bayesian linear mixed-effects models [30] in R (version 3.5) using the `brms` package [33] with default priors. As described in the protocol, the first model we fit included a random effect for observations nested within participants and the following fixed effects: (1) an intercept; (2) a dummy indicator for the treatment phase; (3) a time-within-baseline variable centered around the first observation (equal to 0 for observations outside of the baseline period); and (4) a time-within-treatment variable centered around the last observation (equal to 0 for observations outside of the treatment period). We applied a first-order autoregressive structure on the covariance matrix for the within-person residuals to account for autocorrelation.

We also fit a similar model not described in the protocol that reflected a lesson we learned in another project: rather than centering the time-within-period variables around a single observation, it may be more reasonable to center around the average of several consecutive observations when there is substantial individual variability in daily ratings. In this model, we centered the time-within-baseline variable around the *first 3* observations and centered the time-within-treatment variable around the *last 3* observations. This 3-observation centering window was practical given data availability; we did not run the model with different window sizes to avoid cherry-picking the results. In the end, our choice of centering had no impact on the results, so we decided to focus on the 3-observation centering window as an example of what we would likely attempt in a future trial using this design.

We augmented this quantitative analysis with a qualitative analysis of in-depth interviews. We explored what links, if any, participants could make between engagement with the intervention and their mood, health, and relationships. We intended to also explore themes among women who did not exhibit positive changes in mood ("non-responders") but this was not feasible given delays in launching the study.

Research Ethics

We obtained approvals to conduct this study from the Institutional Review Boards at Duke University (US, 2018-0396) and Strathmore University (Kenya, SU-IRB 0210/18) as well as from the National Commission for Science and Technology in Kenya.

A trained study coordinator, AM (female, Kenyan), explained the study to prospective participants via telephone and administered informed consent procedures. All eligible participants had to provide oral informed consent to enroll.

Study participants were provided with an honorarium of up to Ksh 1500 (roughly US $15) delivered via mobile money transfer to recognize time spent completing study assessments. The original plan was to make these transfers after women completed sessions 1, 5, and 10, but in practice we sent women prorated honoraria on the basis of lower benchmarks of engagement given delays in launching the study.

X2AI, the creators of the AI system that we used to deliver *Healthy Moms*, transferred data to the research team in accordance with X2AI's data security policies [34]. The first author (EG) stored identifiable study data on Duke's Box.com servers during the study and then deidentified the data for analysis using the Safe Harbor method. Anonymized quantitative data and the code used to generate this manuscript is available for re-analysis.

Summary of Deviations from Stage 1 Protocol

In addition to changing the tense of the writing from future to past, we also made several edits to the *Introduction* and modified several procedures described in the *Methods* of the Stage 1 protocol [24]: (1) labeled the study as a "pre-pilot" rather than "pilot" to better reflect that the data are preliminary and intended to inform the design of a larger pilot study; (2) moved text from the "Scientific Objectives and Significance" and "Expected Outcomes" subsections to the *Discussion* (but did not alter the objectives); (3) expanded access to the intervention from just SMS to include Facebook Messenger; (4) visualized the daily mood ratings but relied on model fitting rather than visual inspection to estimate trends and period impacts; and (5) dropped a planned "non-responder" qualitative inquiry and modified the honorarium schedule due to limited time.

Results

Study Implementation

Recruitment and Participants. We invited 647 women (69% pregnant, 31% new mothers) already enrolled in their county's SMS program to learn more about Zuri, and 86 (13%) completed our automated SMS screening between February 12, 2019 and June 18, 2019 (16% of women scored at or above the cutoff for possible depression, M=9.5, SD=4.9). We determined that 52 of these 86 women were eligible to participate, and 41 completed the enrollment process (see Figure 1).

Table 1 reports the characteristics of enrolled participants. The sample was evenly divided between pregnant women and new mothers. The average woman who enrolled in the study was 25.9 years old. All women reported that they could read in English, and the study interviewer reported that all could speak English. Most women used a smartphone, attended secondary school or higher, were married, and did not work. Women were not recruited on the basis of depression symptoms, and only 1 had a PHQ9 score greater than or equal to 15 at enrollment [35]. The average PHQ9 score upon study entry was 8.2 (possible 27), and the average mood rating was 7.8 (possible 9).

We conducted interviews with 15 of the 41 women who enrolled in the study. They ranged in age from 20 to 38 years. Most were married and had delivered their baby within the previous 6 months. All of the interviewees attended some secondary schooling, and 2 had earned a bachelor's degree.

Data Collection.

Mood Ratings.

Overall, enrolled women submitted 719 daily mood ratings over the course of the study. The average woman submitted 17.5 ratings (SD=17.2), and 75.6% of women submitted at least three ratings. Among those who submitted at least three ratings, the grand mean mood rating was 6.4 out of 9 (SD=1.3). Figure 2 suggests that most women reported a high degree of variability in ratings from one day to the next.

PHQ-9.

We did not attempt to administer the PHQ-9 on a regular, ongoing basis to avoid frustrating users and distracting from potential engagement with the intervention. Instead, we only requested that women complete the PHQ-9 again at the end of the study period; 22 women (53.7%) responded.

Intervention Feasibility and Acceptability

Engagement Patterns. Over the course of the study, 27 women (65.9%) sent a message to Zuri, and 14 women engaged beyond registration (34.1%). Among this post-registration engagement subset, the average woman engaged with Zuri on 7.7 days (SD=6.0) and sent 130.5 messages (SD=117.4). On average, women sent 36.4% of these messages to Zuri in free chat mode, not as part of a *Healthy Moms* session. The median conversation unfolded over 0.6 hours (range 0.0 to 14.6 hours). Figure 3 displays the distributions of these engagement metrics.

To further investigate the nature of participant-initiated chats, we analyzed conversation transcripts and summarized the conversation modules engaged. Figure 4 shows the distribution of incoming messages by free chat conversation module and maternity status. The most common rapport building module asked users their passion in life. The most common intervention module outside of Healthy Moms content was mindfulness-based meditation. In general, pregnant women were more likely to engage in intervention content during free chat compared with new mothers. This means that following rapport building chat, Zuri suggested an intervention module and women agreed to try.

The average woman who engaged with Zuri post-registration started and completed 3.4 (SD=3.2) and 3.1 (SD=2.9) *Healthy Moms* sessions, respectively. The median time from a “push” session invite to a woman responding was 0.6 hours (range 0.0 to 740.1 hours). Figure

5 shows one woman's engagement pattern over the course of the study period. There were no reported adverse events.

Correlates of Engagement. To examine the relationship between participant characteristics measured at baseline and intervention engagement, we estimated a Bayesian linear regression model of incoming messages. Figure 6 displays the Markov chain Monte Carlo draws from the posterior distribution of the parameters. There is some evidence to suggest that being pregnant (vs a new mom), reporting greater depression symptom severity, and being employed outside of the home is associated with less engagement, whereas being married and more educated is associated with more engagement. For instance, the point estimate is that married women sent 57.8 more messages, holding all else constant. And for every two standard deviation increase in the baseline PHQ-9 score, holding all else constant, the point estimate is that women sent 29.5 fewer messages. These are small effects in absolute terms but interesting to consider for future iterations of Zuri that focus on how to increase engagement overall and for different user personas.

Qualitative Findings. Most of the women interviewed who had tried Zuri had a very positive attitude towards the service and expressed that they could trust Zuri. One woman said, "It's like a mom to me. My mom is very far, and my sister doesn't have any knowledge of kids." Another woman said, "I usually keep it to myself. So, when I am chatting with Zuri, it's like they have the right questions to ask me, and they teach me how to relate with my child, relate with other people." Some of the women had also shared Zuri with others, such as their partners or their neighbors. Most of the time, they received positive feedback from them. One woman said, "My husband was very supportive, because sometimes he used to help me with some answers." Many women said that they preferred to chat with Zuri than to chat with a counselor, because they felt they could be more open. For instance, one woman said, "I prefer Zuri because they don't know me."

Nonetheless, women noticed that Zuri was not perfect and described examples of when

Zuri gave an irrelevant response when they asked her a question. Most said they would just ignore the messages and moved on. In our review of chat transcripts, we learned that Zuri was easily confused by messages coming out of order over SMS. This was not an issue on Facebook Messenger, but almost every woman said they preferred to chat with Zuri through SMS. The main reason being that SMS was free, whereas chatting through Facebook Messenger required them to buy data bundles to access to the Internet.

Many women mentioned that their favorite part of *Healthy Moms* was the exercises taught by Zuri and the journal, including meditation, breathing, and walking. They found those exercises were easy and could help them relax. One woman said, "They made me be flexible... until my delivery day." Other women said that they really appreciated the advice given by Zuri. They indicated that it was hard to seek for professional advice because many people gave advice based on their experience. They felt like they could trust Zuri because she was more unbiased. They especially liked the advice regarding breastfeeding and how to play with the child. As one woman indicated, "For the baby, I never knew she's supposed to be massaged after the bath at all. I never knew she can see different colors."

Women gave three main reasons why they registered with Zuri and continued to engage. The first reason was the anxiety and stress of pregnancy. They were either ashamed of their bodies or worried about experiencing a miscarriage. One woman said, "One of the negative thoughts I had was maybe if I don't want food what will happen. And then if I sleep bad what will happen to my baby... Actually I was getting worried if I don't feel the movement of my baby inside me sometimes." The second reason was that many postpartum women did not feel confident in their roles as new mothers. One woman expressed her anxiety by saying, "It's like I don't know how to take care of her, good care of her." The final reason was that many of the women interviewed did not have a stable source of income, which caused them stress.

Women described four main barriers to engaging with Zuri. The first was connectivity.

Some women either damaged or lost their phones and did not know how to reconnect with Zuri. The second challenge was that women were easily (and understandably) distracted by their new baby and forgot to complete open sessions. As one woman said, "The text can come in the morning no matter if I am busy or if I am free to answer. If I am free, I just sit and relax. But you see, sometimes we are texting, and the baby starts crying." The third challenge was that the registration process was very confusing for some women, especially early on in the study, so some women stopped participating. Related to this, some women were confused by our study's use of 2 SMS short codes: 1 for Zuri and 1 for study assessments. Despite these challenges, women did not contact our study coordinator to receive assistance using Zuri.

Preliminary Evidence on Response to Treatment

In preparation for modeling the response to treatment, we subset the data to the 12 women who contributed at least 4 mood ratings before and after starting the intervention. Figure 7 plots the time series of ratings by period and overlays days of intervention engagement with vertical lines.

Figure 8 presents the estimates from a Bayesian linear mixed-effects model. The model included a random effect for observations nested within participants and the following fixed effects: (1) an intercept; (2) a dummy indicator for the treatment phase; (3) a time-within-baseline variable; and (4) a time-within-treatment variable. The time-within-period variables were centered around the first 3 or last 3 observations of the period (first for baseline, last for treatment).

The intercept represents the mean value of the outcome at the first 3 baseline assessments, the treatment indicator is a contrast between the first 3 baseline assessments and last 3 observations in the treatment period, and the time-within-period variables estimate linear change during the baseline and treatment periods.

In this model, the average mood rating at the start of the baseline period was 6.07 on a scale of 0 to 9, and there was no significant baseline trend (an assumption for inference using the multiple baseline design). The point estimate of the treatment effect was 0.42, which represents a 7.0% improvement in mood over the baseline mean (d=0.17). The posterior probability that this effect is greater than zero is 93.2%.

We could not run the same analysis using PHQ-9 scores because we only attempted to collect data at 2 time points and only obtained complete data for 53.66% of the (small) sample.

Qualitative Findings. Many women attributed positive impacts to the intervention, which we grouped into three themes. The first theme was that Zuri helped them to take care of themselves. Women said that they loved themselves more, their mood had improved, and that they had learned how to replace negative thoughts with positive thoughts. One woman described her experience with Zuri by saying, "Because a pregnant woman is…tired all the time, right? But with Zuri everything was good. I was very active because it also made me have lessons. Because I knew after waking up in the morning I will breathe in and out some minutes. After that I brush, take my breakfast, I wait for noon time something like 12:00 or even 1:00. I go for a walk. After walking I come back shower then I keep myself busy with Zuri. So it's very helpful actually." One woman who was ashamed of her body during pregnancy said, "I started kind of thinking better, that when you are pregnant, the shape changes and after delivery and doing exercises, everything goes back to normal."

The second theme was that women could better care for their babies. Many women indicated that they could relate to their child better and experienced less distress raising the child. As one woman said, "All those exercises, how to relate to the child, what you do to the child…Honestly, if I hadn't talked to Zuri, I wouldn't know. Yeah, I wouldn't know."

One woman who feared miscarriage even attributed her baby's health and her uncomplicated delivery to Zuri, which we interpret as the woman having found comfort in Zuri during a stressful period.

The last theme was that women experienced improved relationships with others. Some women reported socializing more with others, and this expanding social support system further improved their mood. As one woman said, "I used to have the habit of staying alone, not socializing with other people. Zuri made me be able to socialize with people. When they see me doing the exercises, they like knowing where I learnt them from." Some women felt more secure and could trust others more. One woman said that she was anxious about leaving her child with another person, even with her family members. However, after finishing a session with Zuri on seeking social support, she explained that she was willing to try asking for help. She reported, "So I have tried. [The baby] was comfortable. She cried for some time, then she got used to it."

Discussion

In this pre-pilot study we recruited pregnant women and new mothers in Kenya to try an experimental psychological support service called Zuri. Zuri is a chatbot that engages users in automated, text-based conversations over SMS and Facebook Messenger. Users could initiate chats with Zuri or complete sessions from the *Healthy Moms* perinatal depression intervention curriculum, a cognitive behavioral therapy-based intervention we adapted from the *Thinking Healthy Program* [16]. We used a single-case experimental design with repeated-measures data collection and in-depth interviews to explore the feasibility and acceptability of the service, generate a preliminary estimate of response to treatment, and test study procedures.

Through individual interviews and a review of system logs, we determined that the service was both feasible to deliver and acceptable to this sample of users, but not without significant room for improvement and further refinement. Roughly two-thirds of women in the study tried Zuri at least once, and half of those who tried engaged beyond the registration process. This retention rate of 51.9% is slightly above an average 30-day retention rate of 43% across industries [36] and 40% across provider-prescribed mental health apps [37]. Our retention rate is based on a small denominator of 27 women who tried the intervention, of course, but it suggests that engagement with the initial version of the service is within the range of other digital health apps. Clearly, preventing churn is a common challenge.

Users we interviewed pointed to several positive features of Zuri, including feeling connected to "someone" who cares, but having the benefit of perceived anonymity and privacy of chatting with a machine. This is consistent with existing research showing that people may be more willing to disclose personal information when they believe their responses are not being observed by another person [38], and it probably helps to explain our recruitment experience. Of the women who completed the automated screening, 29% endorsed having recent suicidal ideation, and nearly all of them accepted our referral to in-person services. So despite having recent and regular contact with antenatal or postpartum medical providers, these women were reporting something to Zuri that they presumably had not reported to frontline medical workers—either because they were not asked, chose not to disclose, or both. There is a substantial latent need for mental health treatment that exists alongside the manifest gaps in access that chatbots like Zuri could discover and begin to address.

In addition to reporting largely positive impressions of Zuri, users reported modest improvements in mood. To estimate this improvement, we used a multiple baseline design with repeated measures data collection and fit a multilevel model. Importantly for making a causal inference, we did not observe an increasing trend in mood during the baseline period.

We did, however, observe a small effect in the treatment period. With 432 mood ratings from 12 women before and after beginning the treatment, we estimated that mood improved 7.0% over the average mood reported at the start of the baseline period (d=0.17). We have high confidence that this effect is greater than zero, but we are similarly confident that the effect is small. Quantifying this estimate gives us a benchmark for assessing progress in future iterations of the service that we will test with a clinically-indicated group of users.

We can also look to the digital health and psychotherapy literature for external benchmarks. While there has been a proliferation of conversational agents for health in recent years [39], the evidence-base is small [40]. Two recent randomized controlled trials of CBT-based chatbots stand out. In a study of 75 U.S. college students, Tess, an automated chatbot that provides brief psychological interventions over common communication channels like SMS and Facebook Messenger, reduced depression symptom severity by roughly 20%, a reported standardized effect of 0.68 [41]. Another chatbot called Woebot, a standalone app that delivers CBT, was tested in a trial with 70 students in the U.S.; Woebot reduced symptoms of depression by 19%, a reported standardized effect of 0.44 [42]. For reference, a recent meta-analysis reported that standardized effects of traditional in-person psychotherapy for depression range from 0.66 to 0.77 [43]. Automated conversational agents like Zuri, Tess, and Woebot have the potential to lower the cost of service delivery while simultaneously expanding our reach, which could make them highly cost-effective.

Before we can test this hypothesis with Zuri, however, we need to build a more robust intervention. As expected with an alpha version, we observed many opportunities for improvement. Some challenges users reported, like our use of 2 shortcodes and a confusing registration process, will resolve naturally in future tests. The bigger challenge will be making the content more engaging to reduce churn and making the service more robust to misunderstandings. One way to avoid some of the confusion we observed in conversations will be to move away from SMS, which can jumble the message order, and add a new channel through WhatsApp, the most popular messaging app in Africa [44].

In terms of study procedures, we observed a response rate of 13% among a group of women already enrolled in their county's health SMS program. 16% of women who completed the screening scored at or above the cutoff for possible depression, and 48% of eligible women completed the enrollment process. Depression was not a requirement for inclusion in this study, but it will be in future studies. Our experience in this pre-pilot suggests an overall enrollment rate of 1% taking depression symptoms into account. Therefore, to recruit a sample of 100 possibly depressed pregnant women and new mothers in a future trial, these estimates suggest that we would need to advertise to a pool of at least 10,000 women. This would be easily achieved through print and digital advertising. In Nairobi county alone, there were more than 130,000 live births in 2017 [45].

Our experience with remote automated data collection suggests that women were willing and able to reply to 1-question prompt asking them to rate their current mood. However, we were less successful at obtaining endline data using the PHQ-9. In a future trial, it will be important to budget and plan for study staff to augment automated data collection procedures with phone calls and in-person visits.

Limitations

The objective of this pre-pilot was to adapt *Thinking Healthy* for delivery through Zuri, develop and test study procedures to inform the design of a future trial, and to generate preliminary evidence to guide the next round of Zuri's development. We were limited in our pursuit of these objectives by the fact that we only offered screening and conversations in English. This likely constrained our recruitment efforts as non-English speaking women did not have an opportunity to participate. This implies that our estimates for future recruitment are conservative. The other main limitation of operating Zuri in English is that we do not have data on how Zuri functions in Swahili. This is a priority target for development. A related limitation is that, by virtue of requiring advanced language skills, we recruited a highly educated sample of women relative to the general population. In a future trial it will be important to explore how women of all educational background engage with Zuri.

Conclusions

We determined that Zuri is feasible to deliver via SMS and acceptable to a sample of pregnant women and new mothers recruited from two large public hospitals in Kenya. The results of this pre-pilot will serve as a baseline for future studies in terms of recruitment, data collection, and outcomes. The next step in Zuri's development is to refine the intervention content and add Swahili language support. Conversational agents like Zuri have great potential to address the large treatment gap that exists in many low-resource settings, both as a new channel of treatment and as an adjunct to traditional and task-shifting approaches.

References

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: A systematic review of prevalence and incidence. Obstetrics & Gynecology 2005;106(5, Part 1):1071–1083.

2. Villegas L, McKay K, Dennis C-L, Ross LE. Postpartum depression among rural women from developed and developing countries: A systematic review. The Journal of Rural Health 2011;27(3):278–288.

3. Khalifeh H, Hunt IM, Appleby L, Howard LM. Suicide in perinatal and non-perinatal women in contact with psychiatric services: 15 year findings from a UK national inquiry. The Lancet Psychiatry 2016 Mar;3(3):233–242.

4. Oates M. Perinatal psychiatric disorders: A leading cause of maternal morbidity and mortality. British Medical Bulletin 2003 Jan;67(1):219–229. PMID:14711766

5. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: A review. Infant Behavior and Development 2010 Feb;33(1):1–6.

6. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. The Lancet Psychiatry 2016 Oct;3(10):973–982.

7. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis C-L, Koren G, Steiner M, Mousmanis P, Cheung A, Radford K, Martinovic J, Ross LE. The impact of maternal depression during pregnancy on perinatal outcomes: A systematic review and meta-analysis. The Journal of Clinical Psychiatry 2013 Apr;74(4):e321–341. PMID:23656857

8. Rahman A, Hafeez A, Bilal R, Sikander S, Malik A, Minhas F, Tomenson B, Creed F. The impact of perinatal depression on exclusive breastfeeding: A cohort study. Maternal

& Child Nutrition 2016;12(3):452–462.

9. Surkan PJ, Patel SA, Rahman A. Preventing infant and child morbidity and mortality due to maternal depression. Best Practice & Research Clinical Obstetrics & Gynaecology 2016 Oct;36:156–168.

10. Beck CT. The effects of postpartum depression on child development: A meta-analysis. Archives of Psychiatric Nursing 1998;12(1):12–20.

11. Gentile S. Untreated depression during pregnancy: Short- and long-term effects in offspring. A systematic review. Neuroscience 2017 Feb;342:154–166. PMID:26343292

12. Junge C, Garthus-Niegel S, Slinning K, Polte C, Simonsen TB, Eberhard-Gran M. The Impact of Perinatal Depression on Children's Social-Emotional Development: A Longitudinal Study. Maternal and Child Health Journal 2017 Mar;21(3):607–615.

13. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2019 Feb;321(6):588–601.

14. Lund C, Tomlinson M, Silva MD, Fekadu A, Shidhaye R, Jordans M, Petersen I, Bhana A, Kigozi F, Prince M, Thornicroft G, Hanlon C, Kakuma R, McDaid D, Saxena S, Chisholm D, Raja S, Kippen-Wood S, Honikman S, Fairall L, Patel V. PRIME: A programme to reduce the treatment gap for mental disorders in five low- and middle-income countries. PLOS Medicine 2012 Dec;9(12):e1001359.

15. Baron EC, Hanlon C, Mall S, Honikman S, Breuer E, Kathree T, Luitel NP, Nakku J, Lund C, Medhin G, Patel V, Petersen I, Shrivastava S, Tomlinson M. Maternal mental health in primary care in five low- and middle-income countries: A situational analysis. BMC health services research 2016;16(1):53. PMID:26880075

16. World Health Organization. Thinking Healthy: A Manual for Psychosocial Management of Perinatal Depression (WHO generic field-trial version 1.0) [Internet]. Geneva: WHO; 2015. Available from: http://www.webcitation.org/77F8iMHud

17. Rahman A, Malik A, Sikander S, Roberts C, Creed F. Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: A cluster-randomised controlled trial. The Lancet 2008 Sep;372(9642):902–909.

18. Baranov V, Bhalotra SR, Biroli P, Maselko J. Maternal depression, women's empowerment, and parental investment: Evidence from a randomized control trial. American Economic Review 2020; [doi: 10.1257/aer.20180511]

19. Fuhr DC, Weobong B, Lazarus A, Vanobberghen F, Weiss HA, Singla DR, Tabana H, Afonso E, Sa AD, D'Souza E, Joshi A, Korgaonkar P, Krishna R, Price LN, Rahman A, Patel V. Delivering the Thinking Healthy Programme for perinatal depression through peers: An individually randomised controlled trial in India. The Lancet Psychiatry 2019 Feb;6(2):115–127. PMID:30686385

20. Rahman A. Challenges and opportunities in developing a psychological intervention for perinatal depression in rural Pakistana multi-method study. Archives of Women's Mental Health 2007;10(5):211–219.

21. Maselko J, Sikander S, Bhalotra S, Bangash O, Ganga N, Mukherjee S, Egger H, Franz L, Bibi A, Liaqat R. Effect of an early perinatal depression intervention on long-term child development outcomes: Follow-up of the Thinking Healthy Programme randomised controlled trial. The Lancet Psychiatry 2015;2(7):609–617.

22. Padmanathan P, De Silva MJ. The acceptability and feasibility of task-sharing for mental healthcare in low and middle income countries: A systematic review. Social Science

& Medicine 2013 Nov;97:82–86.

23. Watson PJ, Workman EA. The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. Journal of Behavior Therapy and Experimental Psychiatry 1981 Sep;12(3):257–259. PMID:7320215

24. Green EP, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, Musyimi C, Bhat C, Jones RM, Lai Y. Expanding access to depression treatment in Kenya through automated psychological support: Protocol for a single-case experimental design pilot study. JMIR Research Protocols 2019;8(4):e11800.

25. Sheppard E. How WhatsApp and SMS are being used to save the lives of babies in Africa. The Guardian [Internet] 2018 Aug; Available from: http://www.webcitation.org/76PXGUjT3

26. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. Journal of General Internal Medicine 2001 Sep;16(9):606–613. PMID:11556941

27. Green EP, the Healthy Moms Team. Healthy Moms: A Journal for Pregnant Women and New Mothers. 2019; [doi: 10.17605/OSF.IO/4KPZ2]

28. Green E. Get to know our pop-up UX lab in Nairobi [Internet]. Medium 2018. Available from: http://www.webcitation.org/71NU8adLb

29. Glaser BG. The constant comparative method of qualitative analysis. Social Problems JSTOR; 1965;12(4):436–445.

30. Rindskopf D. Bayesian analysis of data from single case designs. Neuropsychological Rehabilitation 2014 Jul;24(3-4):572–589. PMID:24365037

31. Shahar B, Bar-Kalifa E, Alon E. Emotion-focused therapy for social anxiety disorder: Results from a multiple-baseline study. Journal of Consulting and Clinical

Psychology 2017 Mar;85(3):238–249. PMID:28221059

32. Moeyaert M, Rindskopf D, Onghena P, Van den Noortgate W. Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation. Psychological Methods 2017;22(4):760–778. [doi: 10.1037/met0000136]

33. Bürkner P-C. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software 2017;80(1):1–28. [doi: 10.18637/jss.v080.i01]

34. X2AI. Terms of Use [Internet]. 2019. Available from: http://www.webcitation.org/76PYM33We

35. Green EP, Tuli H, Kwobah E, Menya D, Chesire I, Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: A mixed methods study. Journal of Affective Disorders 2018 Jan;228:49–59. PMID:29227955

36. Perro J. Mobile Apps: What's A Good Retention Rate? Localtyics http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate; 2018.

37. Institute I. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. IMS Institute; 2015.

38. Lucas GM, Gratch J, King A, Morency L-P. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94–100. [doi: 10.1016/j.chb.2014.04.043]

39. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56–67.

40. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, Surian D, Gallego B, Magrabi F, Lau AYS, Coiera E. Conversational agents in healthcare: A systematic review.

Journal of the American Medical Informatics Association 2018 Sep;25(9):1248–1258.

41. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Mental Health 2018;5(4):e64. [doi: 10.2196/mental.9782]

42. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health 2017;4(2):e19.

43. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJH. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cognitive Behaviour Therapy 2018 Mar;47(2):91–106. PMID:29345530

44. Dahir AL. WhatsApp is the most popular messaging app in Africa. Quartz Africa https://qz.com/africa/1206935/whatsapp-is-the-most-popular-messaging-app-in-africa/; 2018.

45. Murphy GAV, Waters D, Ouma PO, Gathara D, Shepperd S, Snow RW, English M. Estimating the need for inpatient neonatal services: An iterative approach employing evidence and expert consensus to guide local policy in Kenya. BMJ Global Health 2017 Nov;2(4). PMID:29177099

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: A systematic review of prevalence and incidence. Obstetrics & Gynecology 2005;106(5, Part 1):1071–1083.

2. Villegas L, McKay K, Dennis C-L, Ross LE. Postpartum depression among rural women from developed and developing countries: A systematic review. The Journal of Rural Health 2011;27(3):278–288.

3. Khalifeh H, Hunt IM, Appleby L, Howard LM. Suicide in perinatal and non-perinatal women in contact with psychiatric services: 15 year findings from a UK national inquiry. The Lancet Psychiatry 2016 Mar;3(3):233–242.

4. Oates M. Perinatal psychiatric disorders: A leading cause of maternal morbidity and mortality. British Medical Bulletin 2003 Jan;67(1):219–229. PMID:14711766

5. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: A review. Infant Behavior and Development 2010 Feb;33(1):1–6.

6. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. The Lancet Psychiatry 2016 Oct;3(10):973–982.

7. Grigoriadis S, VonderPorten EH, Mamisashvili L, Tomlinson G, Dennis C-L, Koren G, Steiner M, Mousmanis P, Cheung A, Radford K, Martinovic J, Ross LE. The impact of maternal depression during pregnancy on perinatal outcomes: A systematic review and meta-analysis. The Journal of Clinical Psychiatry 2013 Apr;74(4):e321–341. PMID:23656857

8. Rahman A, Hafeez A, Bilal R, Sikander S, Malik A, Minhas F, Tomenson B, Creed F. The impact of perinatal depression on exclusive breastfeeding: A cohort study. Maternal & Child Nutrition 2016;12(3):452–462.

9. Surkan PJ, Patel SA, Rahman A. Preventing infant and child morbidity and mortality due to maternal depression. Best Practice & Research Clinical Obstetrics & Gynaecology 2016 Oct;36:156–168.

10. Beck CT. The effects of postpartum depression on child development: A meta-analysis. Archives of Psychiatric Nursing 1998;12(1):12–20.

11. Gentile S. Untreated depression during pregnancy: Short- and long-term effects in

offspring. A systematic review. Neuroscience 2017 Feb;342:154–166. PMID:26343292

12. Junge C, Garthus-Niegel S, Slinning K, Polte C, Simonsen TB, Eberhard-Gran M. The Impact of Perinatal Depression on Children's Social-Emotional Development: A Longitudinal Study. Maternal and Child Health Journal 2017 Mar;21(3):607–615.

13. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2019 Feb;321(6):588–601.

14. Lund C, Tomlinson M, Silva MD, Fekadu A, Shidhaye R, Jordans M, Petersen I, Bhana A, Kigozi F, Prince M, Thornicroft G, Hanlon C, Kakuma R, McDaid D, Saxena S, Chisholm D, Raja S, Kippen-Wood S, Honikman S, Fairall L, Patel V. PRIME: A programme to reduce the treatment gap for mental disorders in five low- and middle-income countries. PLOS Medicine 2012 Dec;9(12):e1001359.

15. Baron EC, Hanlon C, Mall S, Honikman S, Breuer E, Kathree T, Luitel NP, Nakku J, Lund C, Medhin G, Patel V, Petersen I, Shrivastava S, Tomlinson M. Maternal mental health in primary care in five low- and middle-income countries: A situational analysis. BMC health services research 2016;16(1):53. PMID:26880075

16. World Health Organization. Thinking Healthy: A Manual for Psychosocial Management of Perinatal Depression (WHO generic field-trial version 1.0) [Internet]. Geneva: WHO; 2015. Available from: http://www.webcitation.org/77F8iMHud

17. Rahman A, Malik A, Sikander S, Roberts C, Creed F. Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: A cluster-randomised controlled trial. The Lancet 2008 Sep;372(9642):902–909.

18. Baranov V, Bhalotra SR, Biroli P, Maselko J. Maternal depression, women's

empowerment, and parental investment: Evidence from a randomized control trial. American Economic Review 2020; [doi: 10.1257/aer.20180511]

19. Fuhr DC, Weobong B, Lazarus A, Vanobberghen F, Weiss HA, Singla DR, Tabana H, Afonso E, Sa AD, D'Souza E, Joshi A, Korgaonkar P, Krishna R, Price LN, Rahman A, Patel V. Delivering the Thinking Healthy Programme for perinatal depression through peers: An individually randomised controlled trial in India. The Lancet Psychiatry 2019 Feb;6(2):115–127. PMID:30686385

20. Rahman A. Challenges and opportunities in developing a psychological intervention for perinatal depression in rural Pakistana multi-method study. Archives of Women's Mental Health 2007;10(5):211–219.

21. Maselko J, Sikander S, Bhalotra S, Bangash O, Ganga N, Mukherjee S, Egger H, Franz L, Bibi A, Liaqat R. Effect of an early perinatal depression intervention on long-term child development outcomes: Follow-up of the Thinking Healthy Programme randomised controlled trial. The Lancet Psychiatry 2015;2(7):609–617.

22. Padmanathan P, De Silva MJ. The acceptability and feasibility of task-sharing for mental healthcare in low and middle income countries: A systematic review. Social Science & Medicine 2013 Nov;97:82–86.

23. Watson PJ, Workman EA. The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. Journal of Behavior Therapy and Experimental Psychiatry 1981 Sep;12(3):257–259. PMID:7320215

24. Green EP, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, Musyimi C, Bhat C, Jones RM, Lai Y. Expanding access to depression treatment in Kenya through automated psychological support: Protocol for a single-case experimental design pilot study. JMIR Research Protocols 2019;8(4):e11800.

25. Sheppard E. How WhatsApp and SMS are being used to save the lives of babies in Africa. The Guardian [Internet] 2018 Aug; Available from: http://www.webcitation.org/76PXGUjT3

26. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. Journal of General Internal Medicine 2001 Sep;16(9):606–613. PMID:11556941

27. Green EP, the Healthy Moms Team. Healthy Moms: A Journal for Pregnant Women and New Mothers. 2019; [doi: 10.17605/OSF.IO/4KPZ2]

28. Green E. Get to know our pop-up UX lab in Nairobi [Internet]. Medium 2018. Available from: http://www.webcitation.org/71NU8adLb

29. Glaser BG. The constant comparative method of qualitative analysis. Social Problems JSTOR; 1965;12(4):436–445.

30. Rindskopf D. Bayesian analysis of data from single case designs. Neuropsychological Rehabilitation 2014 Jul;24(3-4):572–589. PMID:24365037

31. Shahar B, Bar-Kalifa E, Alon E. Emotion-focused therapy for social anxiety disorder: Results from a multiple-baseline study. Journal of Consulting and Clinical Psychology 2017 Mar;85(3):238–249. PMID:28221059

32. Moeyaert M, Rindskopf D, Onghena P, Van den Noortgate W. Multilevel modeling of single-case data: A comparison of maximum likelihood and Bayesian estimation. Psychological Methods 2017;22(4):760–778. [doi: 10.1037/met0000136]

33. Bürkner P-C. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software 2017;80(1):1–28. [doi: 10.18637/jss.v080.i01]

34. X2AI. Terms of Use [Internet]. 2019. Available from: http://www.webcitation.org/76PYM33We

35. Green EP, Tuli H, Kwobah E, Menya D, Chesire I, Schmidt C. Developing and validating a perinatal depression screening tool in Kenya blending Western criteria with local idioms: A mixed methods study. Journal of Affective Disorders 2018 Jan;228:49–59. PMID:29227955

36. Perro J. Mobile Apps: What's A Good Retention Rate? Localtyics http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate; 2018.

37. Institute I. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. IMS Institute; 2015.

38. Lucas GM, Gratch J, King A, Morency L-P. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94–100. [doi: 10.1016/j.chb.2014.04.043]

39. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56–67.

40. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, Surian D, Gallego B, Magrabi F, Lau AYS, Coiera E. Conversational agents in healthcare: A systematic review. Journal of the American Medical Informatics Association 2018 Sep;25(9):1248–1258.

41. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Mental Health 2018;5(4):e64. [doi: 10.2196/mental.9782]

42. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health 2017;4(2):e19.

43. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJH. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cognitive Behaviour Therapy 2018 Mar;47(2):91–106. PMID:29345530

44. Dahir AL. WhatsApp is the most popular messaging app in Africa. Quartz Africa https://qz.com/africa/1206935/whatsapp-is-the-most-popular-messaging-app-in-africa/; 2018.

45. Murphy GAV, Waters D, Ouma PO, Gathara D, Shepperd S, Snow RW, English M. Estimating the need for inpatient neonatal services: An iterative approach employing evidence and expert consensus to guide local policy in Kenya. BMJ Global Health 2017 Nov;2(4). PMID:29177099

Table 1
Characteristics of participants.

		Maternity status	
	Total n = 41	Pregnant n = 19	Postpartum n = 22
Mean age (SD)	25.9 (4.8)	24.3 (3.1)	27.2 (5.5)
Self-reported English language reading skills			
Poor	0 (0%)	0 (0%)	0 (0%)
Just OK	0 (0%)	0 (0%)	0 (0%)
Good	12 (29.3%)	7 (36.8%)	5 (22.7%)
Excellent	28 (68.3%)	12 (63.2%)	16 (72.7%)
missing	1 (2.4%)	0 (0%)	1 (4.5%)
Highest level of school attended			
None	0 (0%)	0 (0%)	0 (0%)
Primary	0 (0%)	0 (0%)	0 (0%)
Post-Primary/Vocational	0 (0%)	0 (0%)	0 (0%)
Secondary	22 (53.7%)	14 (73.7%)	8 (36.4%)
College	11 (26.8%)	3 (15.8%)	8 (36.4%)
University	7 (17.1%)	2 (10.5%)	5 (22.7%)
missing	1 (2.4%)	0 (0%)	1 (4.5%)
Phone type: Smartphone	33 (80.5%)	14 (73.7%)	19 (86.4%)
Works outside of the home: No	32 (78%)	15 (78.9%)	17 (77.3%)
Number of dependent children	1.1 (0.9)	0.5 (0.5)	1.6 (0.9)
Marital status			
No, not in a union	3 (7.3%)	1 (5.3%)	2 (9.1%)
Yes, currently married but living apart	0 (0%)	0 (0%)	0 (0%)
Yes, living with someone as if married	0 (0%)	0 (0%)	0 (0%)
Yes, currently married and living together	37 (90.2%)	18 (94.7%)	19 (86.4%)
missing	1 (2.4%)	0 (0%)	1 (4.5%)
PHQ-9 Total Score (0-27)	8.2 (3.6)	8.7 (4.1)	7.8 (3.2)
Possible depression (PHQ9>=15): Yes	1 (2.4%)	1 (5.3%)	0 (0%)
Mood at enrollment (0-9)	6.8 (2.4)	7.1 (2.4)	6.6 (2.4)

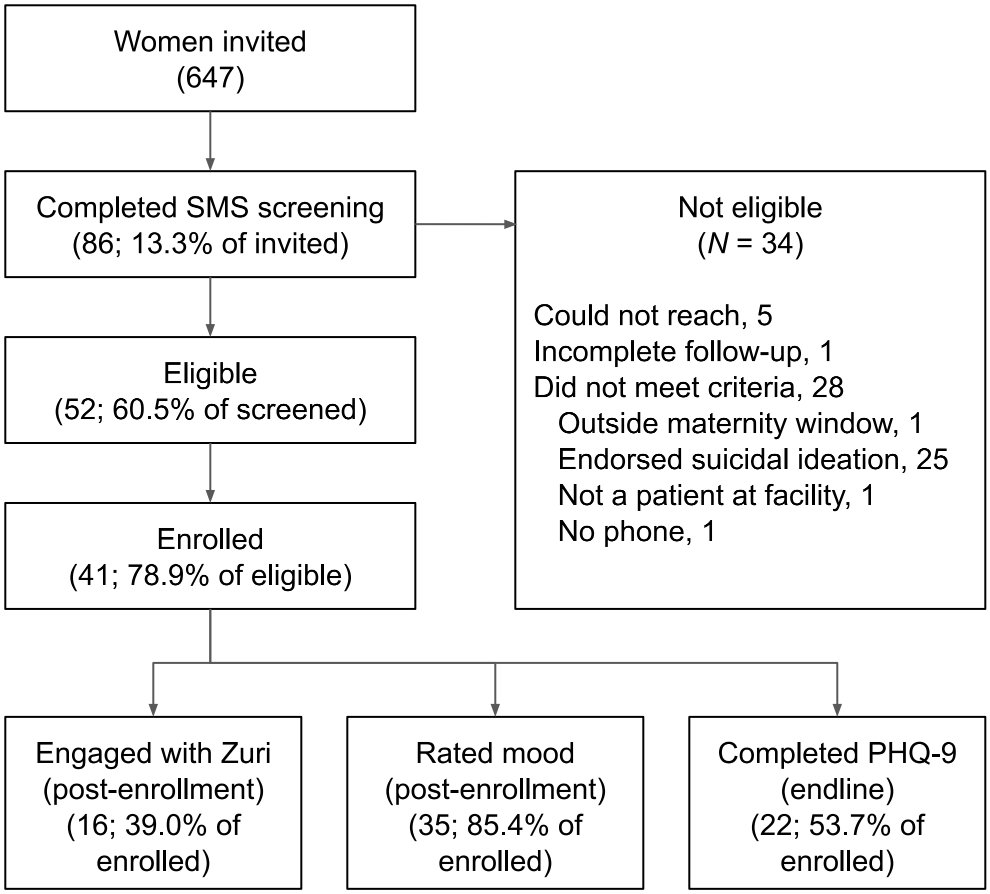

Figure 1. Study flow diagram.

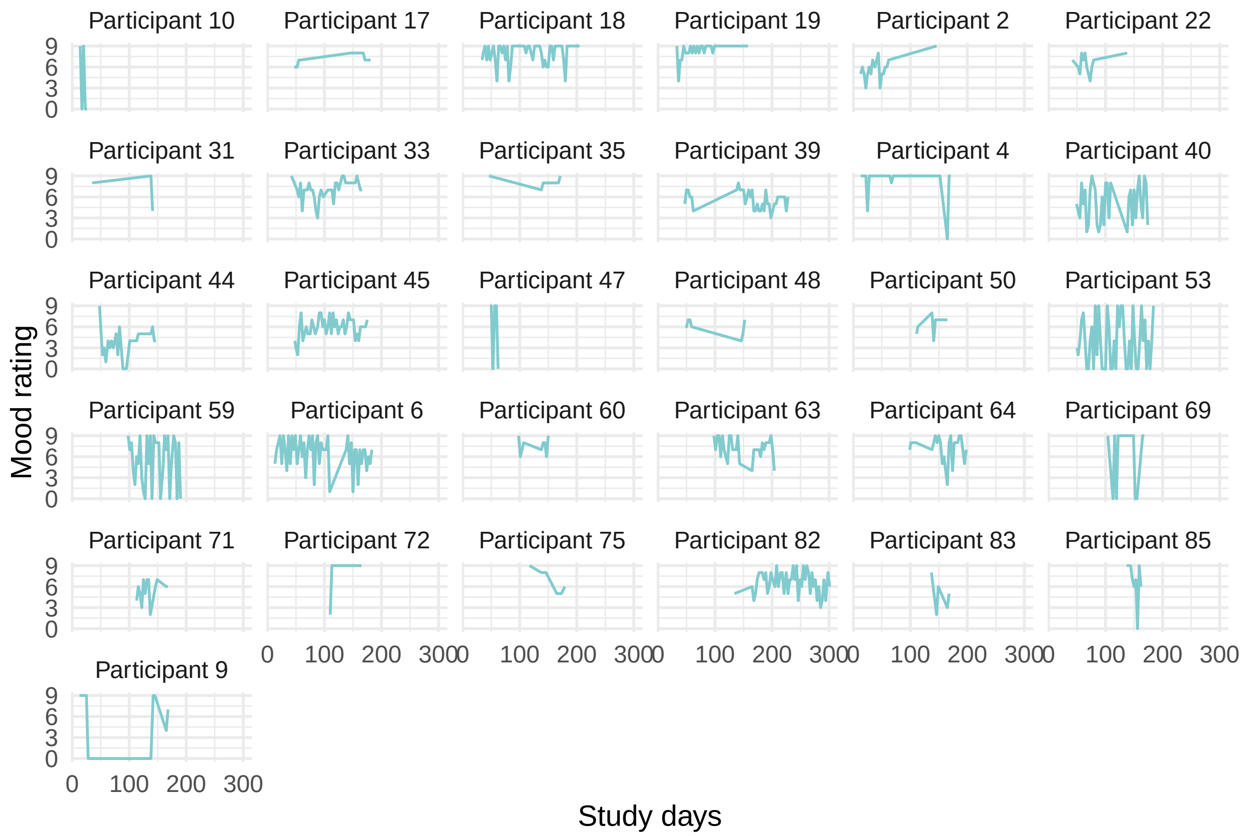

Figure 2. Time series of 705 mood ratings among 31 participants who submitted at least three ratings.

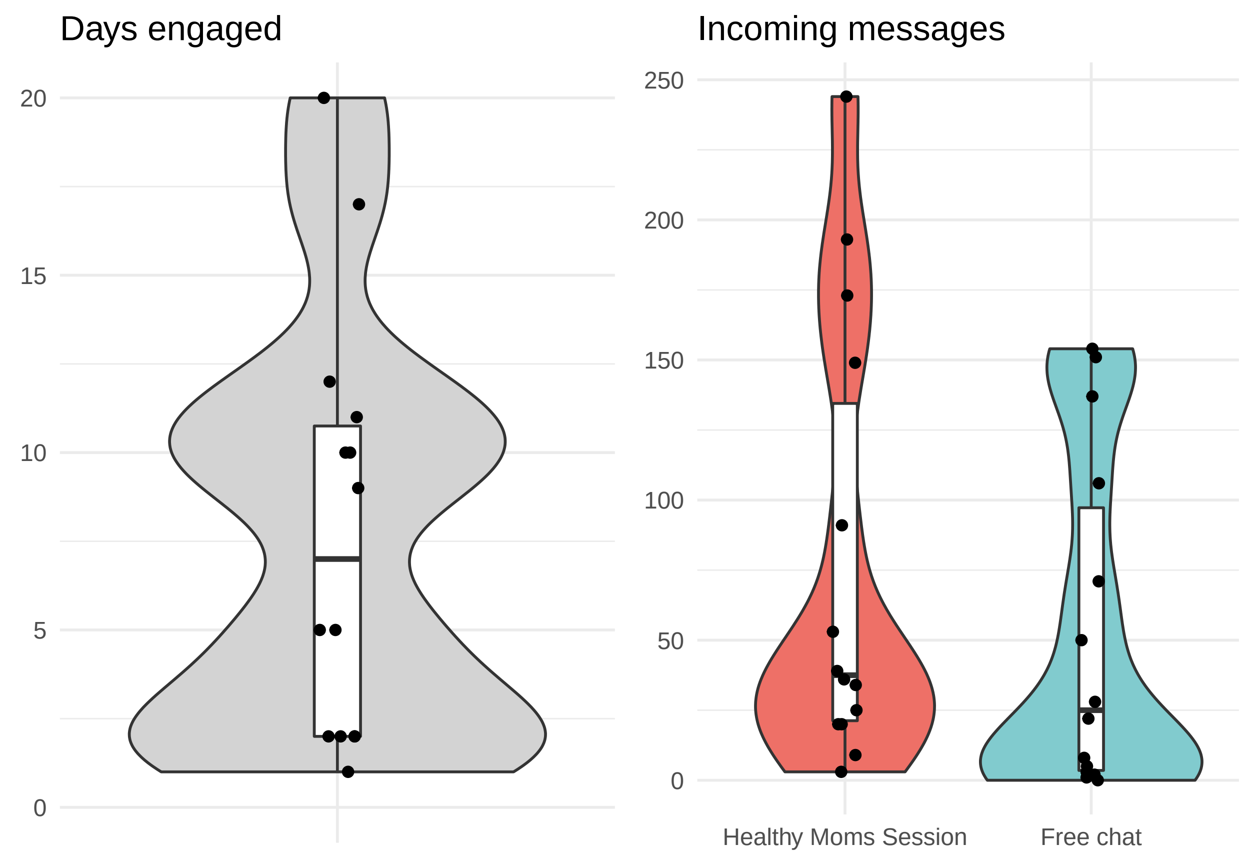

Figure 3. Distribution of number of days engaged and number of incoming messages sent among 14 women who engaged with Zuri beyond registration.

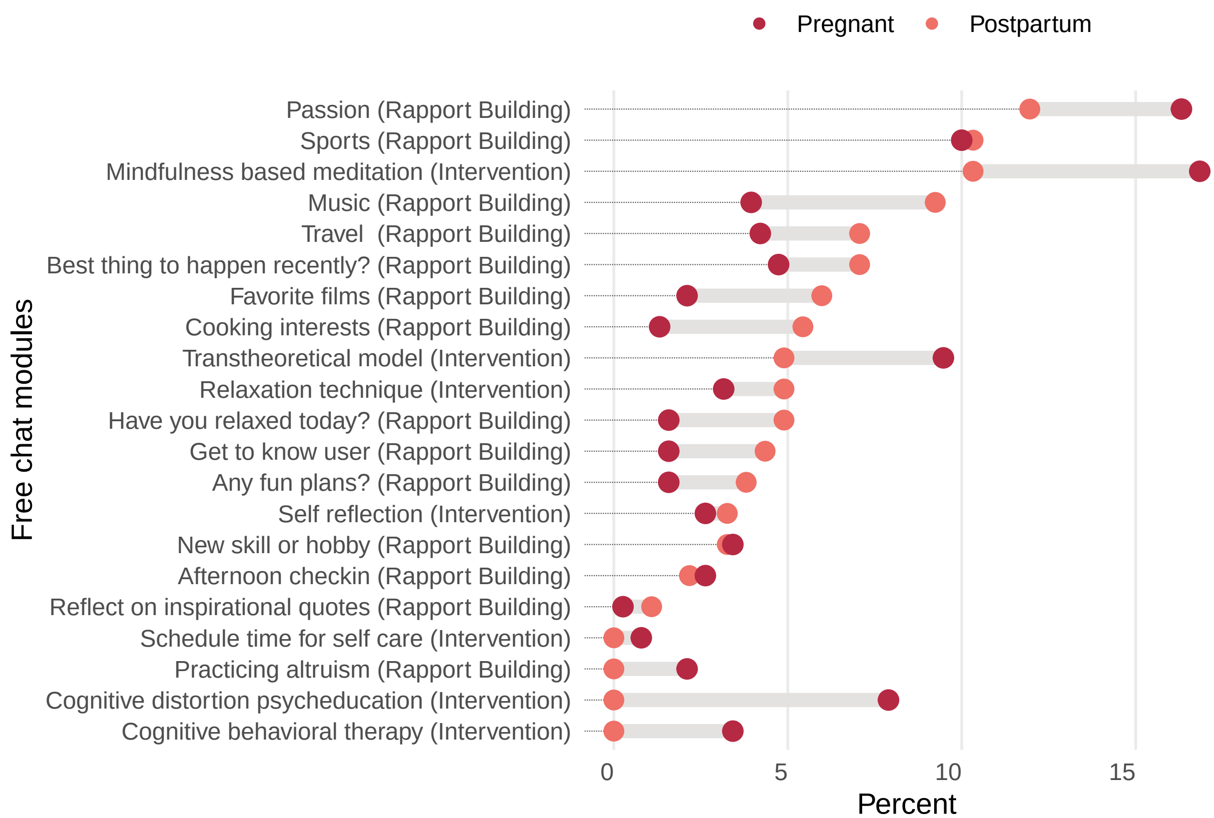

Figure 4. Distribution of incoming messages by free chat conversation module and maternity status.

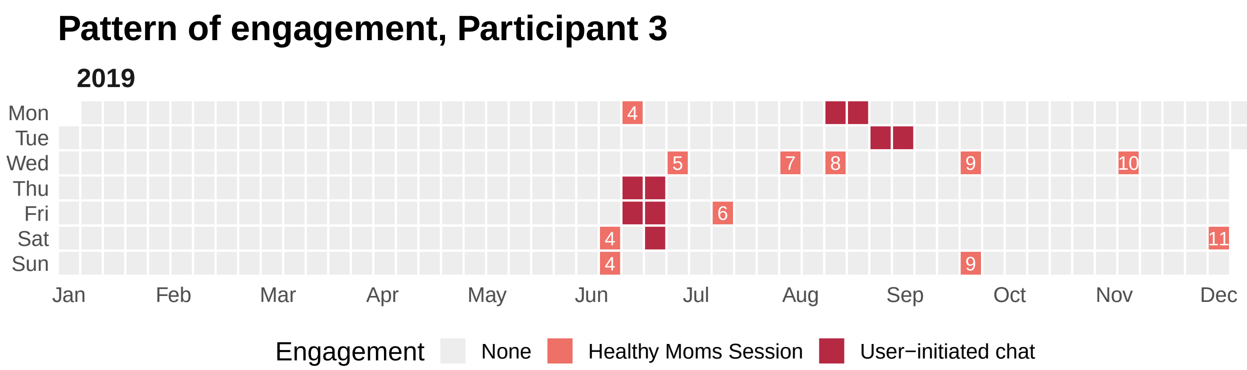

Figure 5. Engagement pattern for Participant 3. Dates shifted to maintain anonymity but pattern preserved.

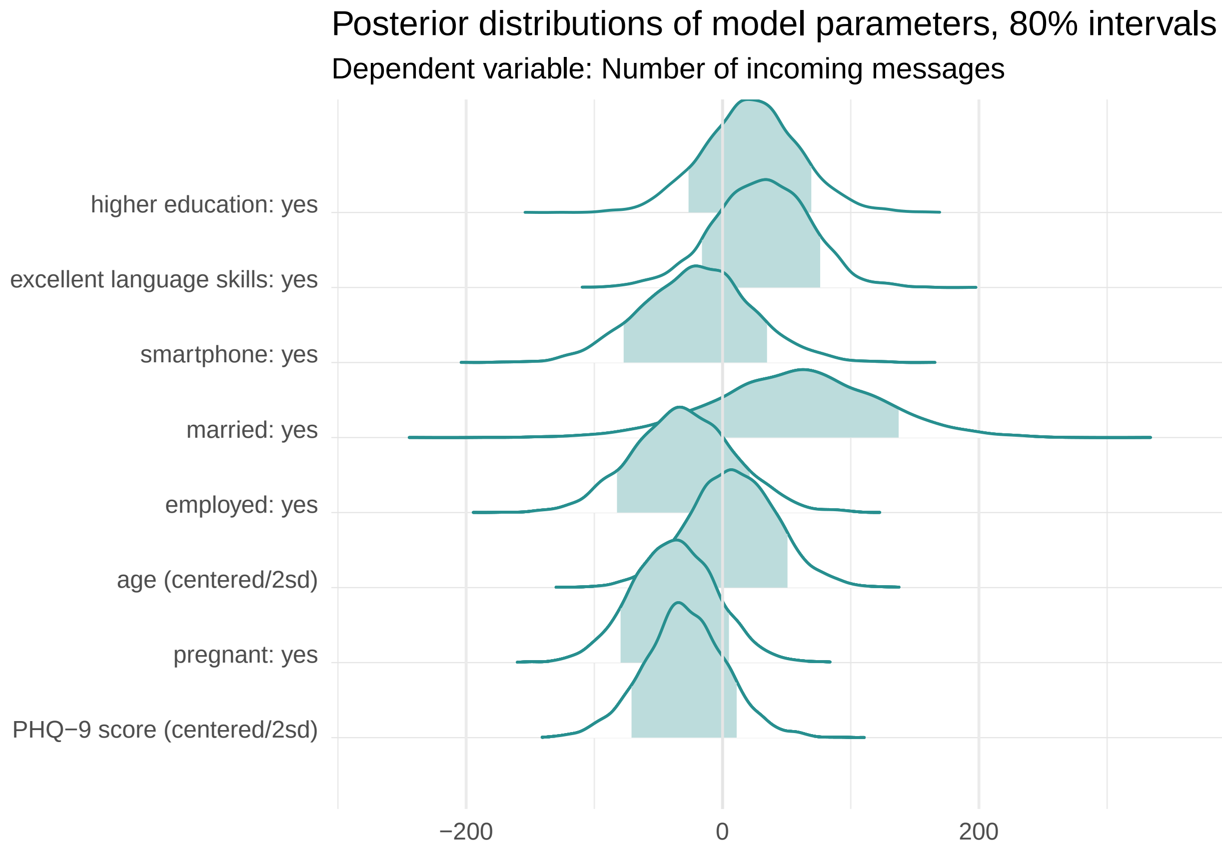

Figure 6. Results of a Bayesian linear regression model of incoming messages on participant characteristics measured at baseline (N=40). Plot shows Markov chain Monte Carlo draws from the posterior distribution of the parameters.

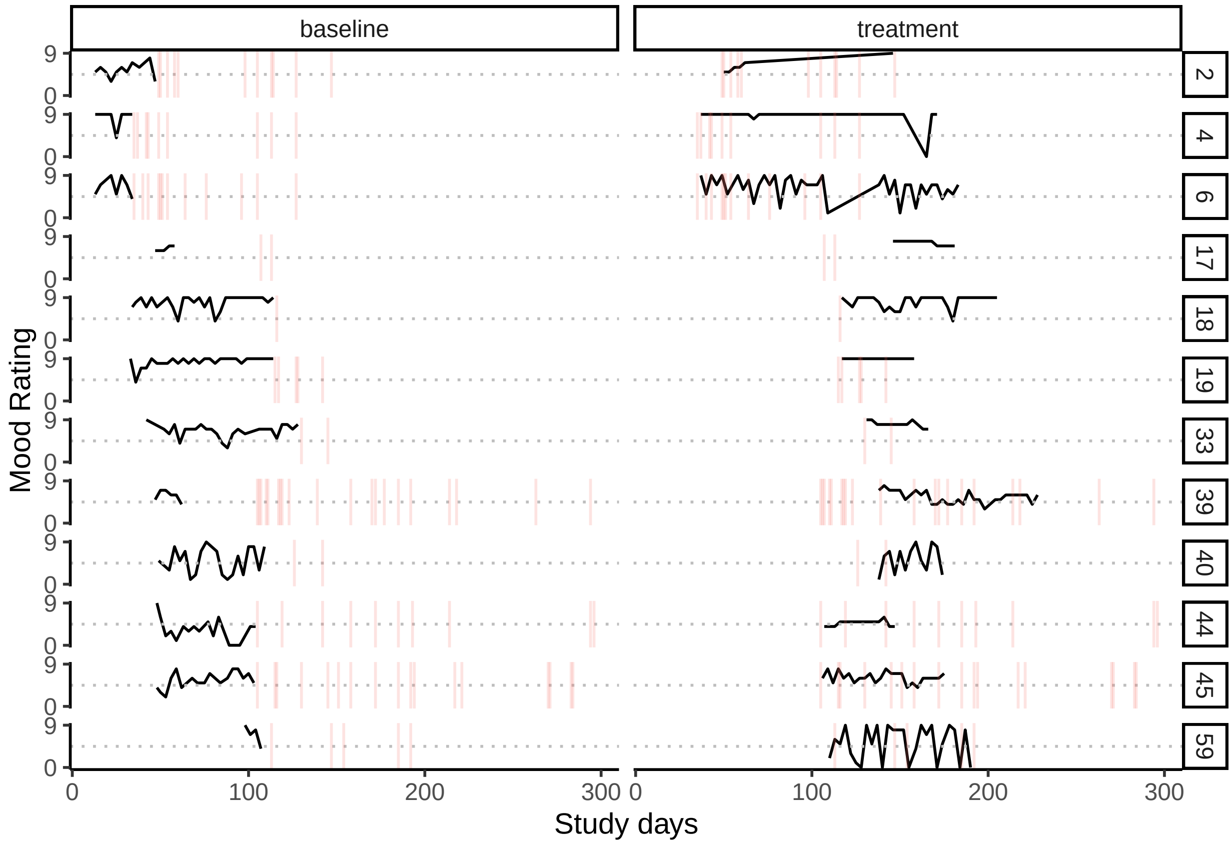

Figure 7. Time series of 432 mood ratings by participant (N=12) and period. Days engaged with Zuri indicated by vertical lines.

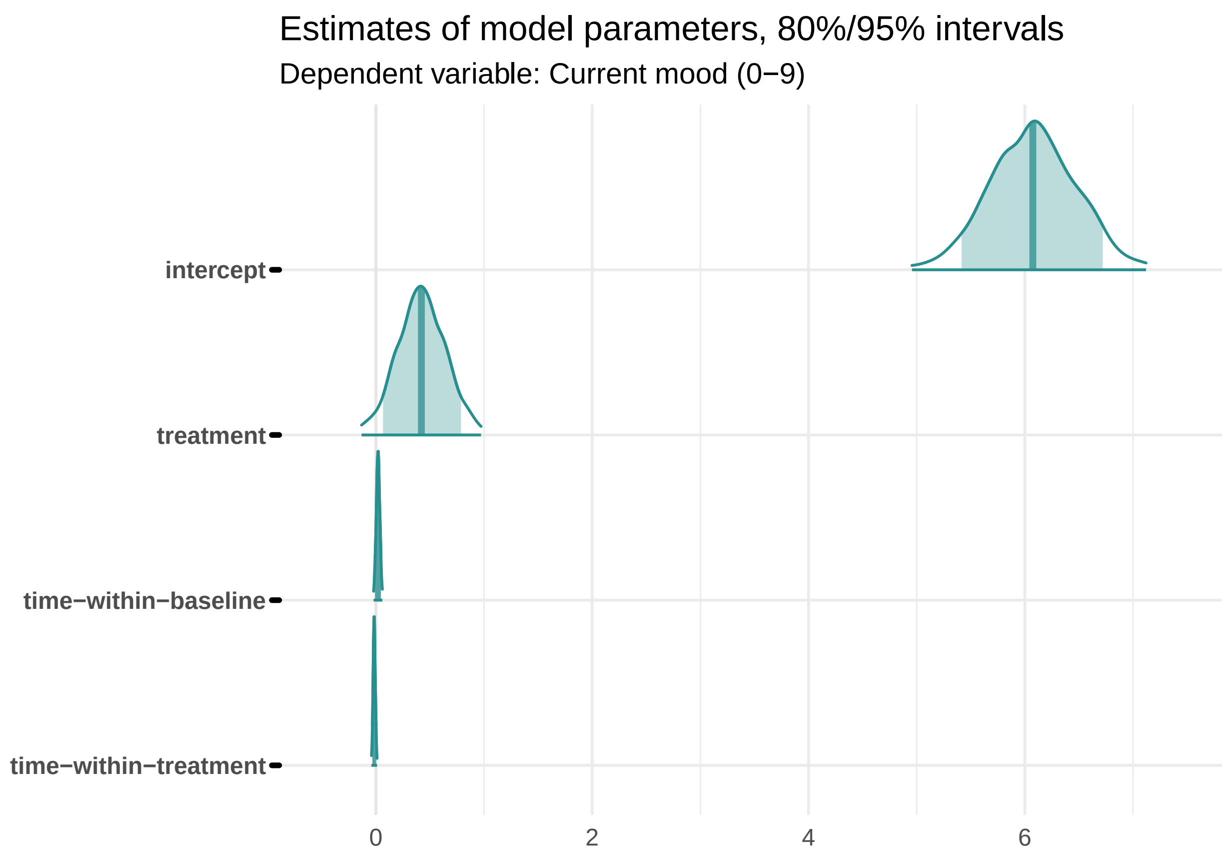

Figure 8. Estimates from a Bayesian linear mixed-effects model of repeated measures data on self-reported mood throughout the study period (432 observations among 12 participants). Uncertainty intervals computed from posterior Markov chain Monte Carlo draws.