CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/134371/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Maeda, Eric, Miyata, Akane, Boivin, Jacky , Nomura, Kyoko, Kumazawa, Yukiyo, Shirasawa, Hiromitsu, Saito, Hidekazu and Terada, Yukihiro 2020. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. Reproductive BioMedicine Online 41 (6) , pp. 113-1143. 10.1016/j.rbmo.2020.09.006

Publishers page: http://dx.doi.org/10.1016/j.rbmo.2020.09.006

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Promoting fertility awareness and preconception health using a chatbot: A randomized controlled trial

Short title:

Chatbot for fertility awareness

Authors:

Eri Maeda [a,*], Akane Miyata [b], Jacky Boivin [c], Kyoko Nomura [a], Yukiyo Kumazawa [d], Hiromitsu Shirasawa [d], Hidekazu Saito [e], Yukihiro Terada [d]

[a] Department of Environmental Health Science and Public Health, Akita University Graduate School of Medicine, Akita 010-8543, Japan

[b] Reproduction Center, Dokkyo Medical University, Saitama 343-8555, Japan

[c] Cardiff Fertility Studies Research Group, School of Psychology, Cardiff University, Cardiff CF10 3AT, United Kingdom

[d] Department of Obstetrics and Gynecology, Akita University Graduate School of Medicine, Akita 010-8543, Japan

[e] Umegaoka Women's Clinic, Tokyo 154-0022, Japan

***Correspondence address.** Eri Maeda, E-mail: erimaeda@med.akita-u.ac.jp

Abstract

Research Question: What are the effects of using a fertility education chatbot (i.e., automatic conversation program) on knowledge, intentions to improve preconception behaviour, and anxiety?

Design: A three-armed, randomized, controlled trial was conducted using an online social research panel. Participants included 927 women aged 20–34 years who were randomly allocated to one of three groups: a fertility education chatbot (intervention group, IG), a document about fertility and preconception health (control group 1, CG1), or a document about an irrelevant topic (control group 2, CG2). Participants' scores on the Cardiff Fertility Knowledge Scale and the State-Trait Anxiety Inventory, their intentions to optimise preconception behaviours (e.g., taking folic acid), and the free-text feedback provided by chatbot users were assessed.

Results: A repeated-measures analysis of variance showed significant fertility knowledge gains after the intervention in the IG (+9.1 points) and CG1 (+14.9 points) but no significant change in CG2 (+1.1 points). Post-test increases in the intentions to optimise behaviours were significantly higher in the IG than in CG2 and were similar to those in CG1. Post-test state anxiety scores were significantly lower in the IG than in CG1 and CG2. User feedbacks about the chatbot suggested technical limitations (e.g., low comprehension of users' words) and pros and cons of using the chatbot (e.g., convenient versus coldness).

Conclusions: Providing fertility education using a chatbot improved fertility knowledge and intentions to optimise preconception behaviour without increasing anxiety, but the improvement in knowledge was small. Further technical development and exploration of personal affinity for technology is required.

Keywords: fertility awareness, education, preconception, chatbot, digital technology

Introduction

Fertility awareness is of growing interest and importance in the world (Zegers-Hochschild et al., 2017). Many people postpone parenthood due to career, education, relationship, and financial issues (Mills et al., 2011); as a result, people sometimes face biological barriers to achieving a desired family size (Habbema et al., 2015). In addition to choices, an increased incidence of non-communicable diseases such as obesity, diabetes, and thyroid disorders (Broughton and Moley, 2017; Thong et al., 2020; Krassas et al., 2010) has caused more women of reproductive age to experience subfertility. In such contexts, fertility education is provided to reproductive-aged people in the community, schools, and health care facilities using various tools, such as brochures, online information, theatre, and educational videos (Daniluk and Koert, 2013; Hvidman et al., 2014; Hammarberg et al., 2013; Boivin et al., 2018a; Harper et al., 2019). These educational interventions improve fertility awareness, both in the short term (Daniluk and Koert, 2015; Wojcieszek and Thompson., 2013; Maeda et al., 2016) and even two years after exposure (Maeda et al., 2018). However, interest in future pregnancy and fertility education is often limited, as is the ability to integrate fertility information into everyday life (Boivin et al., 2018b; Maeda et al., 2018). It is therefore necessary to continue developing strategies to encourage people to participate in their fertility, particularly strategies that can be delivered efficiently to large populations.

Maintaining good preconception health helps ensure successful pregnancies, healthy babies, and good health in the current and next generation (World Health Organization, 2012; Stephenson et al., 2018). Preconception health promotion encourages all reproductive-aged people, irrespective of their current childbearing intentions, to achieve optimal health and wellness, thus ensuring good health for them and any children they may have (Verbiest et al., 2016). Preconception care can include reproductive life plan (RLP) counselling; provision of family planning and contraception; guidance about nutrition, immunizations, infection control, and treatment and monitoring of chronic medical conditions; and information about exposure-related lifestyle choices, such as tobacco and alcohol use and substance abuse (Malnory and Johnson, 2010; Jack et al., 2008).

In Japan, where the total fertility rate is low (1.42 in 2018) and the parental age at first birth is high (30.7 and 32.8 years for women and men, respectively, in 2018), awareness of preconception health seems to be as low as awareness of fertility (Maeda et al., 2015). For example, the main contraceptive method in Japan is condoms (83%), which have a typical use failure rate that is reported to be much higher than that of hormonal methods (13% versus 0.01 to 7%) (Trussell et al., 2018). In addition, very few women (3%) take oral contraceptive pills, despite the obvious reproductive benefits (Yoshida et al., 2016). Once pregnant, only 8% of women in Japan use folic acid supplementation adequately (Ishikawa et al., 2018). Evidence suggests that mental models of pregnancy interfere with preconception health practices. For example, folic acid is believed to not be needed because of perceptions that the good health of the mother protects the pregnancy from threat or that pregnancy has evolved to be naturally robust or immune to risk (Fulford et al. 2014). Promoting knowledge of and involvement with preconception health tackles misconceptions arising from mental models, and it thus seems to be as essential as fertility education among people of reproductive age.

Novel digital technology can be used to deliver low-cost health promotion initiatives at the population level, particularly among those of reproductive age. These digital natives include Millennials born in 1980–1994 and iGen born in 1995 or later (Twenge, 2017). Indeed, mobile health apps, such as *Smarter Pregnancy* (van Dijk et al., 2017) and *Infotility* (Zelkowitz et al., 2019), and virtual animated characters, such as Gabby (Jack et al., 2015), have shown promising results for improving preconception health. The chatting robot, or "chatbot," may also be useful in this context. A chatbot is an information and communication tool that uses natural language processing to interact with users automatically (Schmidlen et al., 2019). The chatbot can be programmed with scripts that provide tailored information to users, although most of these scripts are limited to predetermined scenarios. Chatbots used in customer services and banking industries have also been applied in health care contexts to provide education about sex, drugs, and alcohol (Crutzen et al., 2011) and to screen patients for sexually transmitted infections (Kobori et al., 2018). Thus, although still in its developmental phase, chatbot technology could be a promising strategy for promoting fertility awareness and preconception care.

Previous research suggests that interaction and learning with a virtual agent may mitigate negative emotions; people sometimes feel more comfortable sharing sensitive information with computers, which they perceive to be safer confidantes than other people (Lucas et al., 2014; Palanica et al., 2019). Also, Stein and Brooks (2017) reported that "compassionate" care provided by a chatbot facilitates behavioural changes and weight loss among overweight and obese participants. Previously, our randomized controlled trial showed that fertility education using online brochures improves fertility knowledge, but it increases anxiety among people who want to have a child (Maeda et al., 2016). Given that fertility information often involves private lifestyle information, people may feel more comfortable receiving counselling and information from a new technology than from conventional methods (e.g., brochures).

The aim of the present study was to evaluate whether a chatbot that provides fertility and preconception health education changes the knowledge levels, health-related intentions, and psychological states among reproductive-aged users. We specifically targeted women aged between 20 and 34 years who were assumed to need correct fertility information and to be familiar with digital technology. We randomised reproductive-aged participants into one of three groups: an intervention group (IG), which interacted with an educational chatbot designed to provide fertility and preconception health; a control group (CG1), which received a PDF document about fertility and preconception health; or another control group (CG2), which received a PDF document about an irrelevant topic, the national pension system. We hypothesised that people in the IG would demonstrate a greater increase in knowledge and intentions to optimise preconception lifestyles than those in the control groups. In addition, we hypothesised that people in the IG would show less anxiety than those in the control groups.

Materials and Methods

We conducted a three-armed (one intervention and two control groups), randomized, open-label, controlled trial in March 2019: Trial registration number: UMIN Clinical Trials Registry (UMIN000035736). Participants were randomly assigned to one of three educational materials.

Ethical approval

The ethics committee at Akita University Graduate School of Medicine approved the study protocol on March 29, 2018 (no. 1918).

Participants

Participants were recruited via an online social research panel. Inclusion criteria were being a woman aged 20 to 34 years and hoping to have children (or more children) now or in the future, regardless of any current effort or plan to achieve pregnancy. We excluded women who were currently pregnant. Medical professionals were excluded from recruitment. By default, advertising professionals were excluded from the online social research panel according to the market research company procedures (see “Procedures”). Only those who voluntarily agreed to spend about one hour learning the assigned material were invited to participate in the survey.

Procedures

An online market research company (Macromill, Tokyo, Japan), which has a nationwide social research panel of more than 1 million registrants, sent a pre-screening questionnaire regarding the inclusion criteria to 196,195 randomly selected female registrants aged 20–34 years. Of the 10,000 women who responded to the screening questions, 2,524 were eligible. Among the 1,813 who were randomly selected from the eligible respondents and received recruitment emails, 927 completed the survey (51.1% participation rate among eligible invitees). Participants were then randomized to one of the three previously described groups (for each group, n = 309) using a computerized central allocation system (ScreeningMacro, Macromill, Japan). Participants did not learn of their group assignment until they completed the post-test survey. Figure 1 illustrates the participant selection and randomization process.

All study materials were presented online using Airs software (Macromill). After completing the pre-test survey, participants received instructions for their assigned group. Participants in the IG were instructed to go to a website and chat with the online chatbot. Participants in CG1 and CG2 were instructed to visit a website and read the entire online brochure at the respective site. Then, participants were asked to close the study website and complete a post-test survey. Those who completed the survey were given a coupon, which was consistent with the market research company’s procedures (usually less than 1 Euro). All procedures were conducted from March 13, 2019, to March 22, 2019.

Educational materials

We developed a scripted chatbot for the IG (Figure 2). Scripted chatbots generally involve a predetermined scenario wherein the chatbot responds to the user's input with appropriate, pre-determined information. For our predetermined scenario, we programmed the chatbot to start by asking questions we adapted from RLP counselling and education (Malnory and Johnson, 2010). Specifically, the chatbot's information was excerpted from an educational booklet for general readers by the Japan Society of Obstetrics and Gynecology (2018). Topics included factors with significant impacts on fertility and preconception health: normal and abnormal menstruation; timing of sex to increase the likelihood of pregnancy; infertility (definition, prevalence, causes among men and women, and age-related declines); contraception; abortion; sexually transmitted and other types of infections; common reproductive diseases in young women (fibroids and endometriosis); chronic diseases (e.g., depression and diabetes); other diseases (breast and cervical cancers); appropriate body weight for a safe pregnancy; harmful lifestyle choices, such as smoking, alcohol, and illegal drug consumption; vaccinations; domestic violence; and sexual diversity. The contents and text expressions were simplified and summarized to accommodate the chatting style.

To design the chatbot conversations in line with the RLP counselling style (Stern et al., 2013), we consulted several educational sources, including Habbema et al. (2015) and *A Guide to Fertility* by Boivin (2018a; Cardiff Fertility Studies Research Group, 2016). We drew a flowchart with 8,931 characters in Japanese. The expected conversations were implemented using Google Cloud's Dialogflow, a natural language processing engine. In addition, we appended as many potential phrases and keywords as possible from users and chatbot responses. Prototypes were repeatedly tested and refined, first internally and then by a small group of university students and colleagues at collaborative companies, until the chatbot development team comprising researchers, information technology experts, and designers was satisfied with the response functions and the quality.

Participants in CG1 were provided with a PDF document containing the same fertility and preconception health information as in the chatbot (Japan Society of Obstetrics and Gynecology, 2018). The PDF comprised 43 pages and 42,070 Japanese characters. Participants in CG2 were provided with a PDF document containing information about the national pension system (a topic unrelated to fertility education), which was excerpted from the Ministry of Health, Labour, and Welfare (2017a) website. The PDF comprised 34 pages and 26,233 Japanese characters. We conducted a pilot survey with a small group of our colleagues to ensure that the PDFs and questionnaire were understandable.

Measures

Fertility knowledge

We used the Japanese version of the Cardiff Fertility Knowledge Scale (CFKS-J) (Bunting et al., 2013; Maeda et al., 2015) to assess fertility knowledge on the pre- and post-test surveys as a primary outcome. The CFKS-J uses 13 items to measure knowledge about fertility facts, risks, and myths. All items were rated as ‘true’, ‘false’, or ‘do not know’. A correct answer received one point, and an incorrect or ‘do not know’ answer received zero points. Scores were reported as the percentage of correct answers (0% to 100%). The internal consistency coefficient alpha of the CFKS-J was 0.74, and the scale had a one-factor structure (Maeda et al., 2015).

Preconception health status, behaviour, and intention to change

For the pre-test survey, participants completed a questionnaire about the following health status items: weight (kg), height (cm), severe period pain (yes/no), and presence of abnormal menstrual cycle length (less than 21 days, more than 35 days, irregular, or amenorrhoea) without oral contraceptives (Bunting and Boivin, 2010). Participants also reported on the following behaviour related to their fertility and preconception health; the answers in bold letter were assumed to be preconception behaviours: 1) current smoker (yes/**no**), 2) proactive intake of folic acid supplement or enriched food (**yes**/no), 3) vaccination against HPV infection (**yes**/no), 4) previous cervical check-ups (**yes**/no), and 5) having a primary obstetrics and gynaecology (ob/gyn) doctor (**yes**/no). In Japan, ob/gyn specialists, instead of general physicians, address all primary care for ob/gyn diseases. In addition, we asked participants about 6) currently trying to get pregnant (**yes**/no). For those who were not currently trying to get pregnant, we asked about 7) their use of oral contraceptives (**yes**/no) and 8) other contraceptive methods (**always yes**/no).

In the pre- and post-test surveys, participants who did not exhibit any of the eight preconception behaviours listed above were asked to score their intention to change each behaviour using a three-point scale: ‘preparation’ (i.e., ready to take action), ‘contemplation’ (i.e., interested in changing behaviour but still ambivalent), and ‘precontemplation’ (i.e., not interested in the behaviour). These answers were based on the transtheoretical model before action (Prochaska and DiClemente, 1983).

Post-test psychological assessment

The psychological assessment was administered once during the post-test survey. We used the Japanese version of the State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1970; Nakazato and Mizuguchi, 1982). The STAI uses a 4-point Likert scale (range of 20–80) to measure 20 state-anxiety items (STAI-S), which indicate the current anxiety level, and 20 trait-anxiety items (STAI-T), which indicate the characteristic (trait) anxiety level. Higher scores indicate greater anxiety. The Japanese version of the STAI-S has shown high internal consistency (coefficient alpha = 0.92), and the STAI-T has shown a test-retest reliability of 0.76 for 1 hour later and 0.71 for 3 months later (Nakazato and Mizuguchi, 1982).

Sociodemographic factors

The online market research company provided participant ages. Participants also reported their annual household incomes, university education (yes/no), current marriage status (yes/no), and whether they had a child (yes/no). Annual household income was categorized into four groups: low (< 4 million Japanese Yen), moderate (4–5 million Yen), high (≥6 million Yen), and unknown. At the time of the study, 1 US Dollar = 110 Japanese Yen.

Text analysis

We analysed the free-text feedback qualitatively. Two researchers (EM and AM) separately interpreted, classified, and tallied feedback items by topic. First, each researcher reviewed respondents' feedback and divided the comments into individual, single-meaning text fragments. Second, each researcher grouped similar text fragments together. Both researchers then discussed the shared meanings of each sorted group and classified them into the broadest, but still meaningful, categories. To ensure rigor and consistency of interpretation of the feedback, the researchers discussed any disagreements and reached consensus on all classifications.

Statistical analyses

We estimated the sample size of each group (n = 309) based on the assumption that the mean post-test knowledge scores for the IG and CG1 would increase by 70 ± 23 and 64 ± 23 percent correct scores, respectively, according to results from previous studies (Maeda et al., 2016; Bunting et al., 2013), with 90% power and a significance level of 5%.

We performed all analyses on an intention-to-treat basis. We compared sociodemographic factors, preconception health status, and behaviour between the groups using chi-square tests, one-way analysis of variance (ANOVA), Kruskal-Wallis tests, and post hoc Bonferroni multiple corrections according to the type and distribution of variables. To determine the knowledge difference between groups and over time (pre-test and post-test), we performed a repeated-measures, mixed-factorial, between-within ANOVA using conservative F-tests (Greenhouse-Geisser correction) for the main effect of time and for interactions between groups (IG, CG1, and CG2) and times (pre-test and post-test). Simple effects were used as follow-up tests. To explore between-group differences in pre-test to post-test changes in intention to adopt preconception behaviour, we compared pre–post differences for each person between groups using a nonparametric, pairwise, multiple-comparison procedure following Kruskal–Wallis tests, or Dunn's test (Dinno, 2015). All analyses were performed using STATA14-MP (StataCorp LP, College Station, TX, USA). A two-sided P-value of <0.05 was considered statistically significant.

Results

Background characteristics and group equivalence

Table 1 shows the demographic characteristics of the 927 participants. Participants were about 29 years old, and more than 60% had a university education, which was higher than the national university enrolment ratio of 43% among female high school graduates in 2008 (Ministry of Education, Culture, Sports, Science and Technology, 2008). Less than half were married, and most had no children. Baseline sociodemographic status was well-balanced between groups. Regarding preconception health status and behaviour, about 20% were underweight and 10% were overweight or obese, which is similar to national statistics of 22% and 6%, respectively, among women in their 20s (Ministry of Health, Labour, and Welfare, 2017b). More than half reported severe period pains or abnormal menstrual cycles, but less than 30% had a primary ob/gyn. Fewer than 20% of women reported taking proactive folic acid or receiving an HPV vaccination. Among those who were not currently trying to get pregnant (81% of the participants), 67% stated that they always used contraceptive methods and 9% reported using oral contraceptives, which was a higher oral contraceptive use rate than that of a nationally representative sample (3%) (Yoshida et al., 2016). Baseline health status and behaviours were well-balanced between groups, although the proportion of participants having severe period pain was significantly lower in CG1 than in the IG (Bonferroni adjusted $P = 0.03$).

Effect of the intervention on outcomes

We recorded 574 chatbot sessions, which had an average length of 8 minutes. Because the chatbot was located on the private website during the survey period, multiple sessions were recorded per participant.

Fertility knowledge

The percentages of correct scores on the pre-test CFKS-J were similar between groups (mean ± SD was 59.5 ± 22.7 for the IG, 61.5 ± 20.6 for the CG1, and 60.9 ± 21.9 for the CG2; $P = 0.53$), as shown in Figure 3A. A repeated-measures ANOVA of the scores on the CFKS-J showed a significant interaction between group and time (F [2, 924] = 51.1, $P < 0.001$). Simple effects of time for each group showed that knowledge improved over time in the IG (+9.1 points, 15% gain, $P < 0.001$), CG1 (+14.9 points, 24% gain, $P < 0.001$), and CG2 (+1.1 points, 2% gain, $P = 0.24$). The post-test CFKS-J score for the IG (68.7 ± 23.0) was 7.7 points lower than that of CG1 (76.4 ± 18.4, $P < 0.001$) and 6.7 points higher than that of CG2 (62.0 ± 23.6, $P < 0.001$).

Intention to change preconception behaviour

Participants who did not exhibit preventive behaviours on the pre-test survey scored their intentions to change each behaviour before and after exposure to information in their respective groups. As shown in Table 2, the pre-test to post-test increase of intentions to take folic acid, to receive HPV vaccination, to obtain a primary ob/gyn, to take oral contraceptives, and to try to get pregnant were significantly higher in the IG than in CG2. Compared with CG1, the increase in the intention to take folic acid was significantly higher in the IG, whereas the intention to take oral contraceptives was significantly lower in the IG. Even after considering the possible alpha inflations for the family comparisons (i.e., eight behaviours) by applying additional Bonferroni corrections, the results did not change except for the intention to try to get pregnant in the IG and CG2.

Post-test psychological assessment

Post-test state anxiety scores on the STAI (mean ± SD) were significantly lower (less anxiety) in the IG (43.2 ± 9.5) than in CG1 (47.5 ± 9.5) and CG2 (46.2 ± 9.0), all $P < 0.001$. No difference in post-test trait anxiety scores existed between groups, indicating that differences in state anxiety was not due to underlying differences in personality traits between groups (Figure 3B).

Feedback from chatbot users

Of the 309 participants in the IG, 278 provided text feedback after the intervention (52 Japanese characters, on average). Three topics were identified among the 275 specific comments, including technical problems, pros and cons of using the chatbot, and experiences learning about fertility and preconception health. Twenty-eight participants (10.2%) reported technical problems (e.g., *"It froze up soon"* and *"I could not chat at all*"), and 77 (28.0%) mentioned low comprehension of the chatbot (e.g., *"I rephrased some words when the chatbot did not* *understand"* and *"There were problems of misunderstanding*"). Fifteen (5.5%) comments noted that the chatbot operation was too slow, and another 14 (5.1%) reported that the information was displayed too quickly.

Regarding the pros and cons of using the chatbot, 96 (34.9%) mentioned pros and 33 (12.0%) mentioned cons. Benefits cited included that the experience was "*fun*", "*interesting*", "*easy*", "*convenient*", "*casual*", and "*did not make users feel embarrassment or shyness during chatting about reproductive health … because it is a chatbot*". Among these positive comments, 28 (10.2%) indicated that learning through chatting could promote understanding more than just reading. They mentioned that "*chatting style could lead to better understanding*" and that it was "*easier to understand, compared to ordinary learning accompanied by reading long sentences*". On the other hand, 15 (5.5%) noted that chatting was burdensome and unnecessary. They mentioned that "*reading good websites would be more impressive and readable than using chatbot*" and that "*typing is burdensome*". A lack of humanity or empathy (e.g., "*robotism*", "*coldness*", or "*one-way interaction*") was mentioned in 17 (6.2%) comments. Users "*felt like* [I was] *being replied* [to] *automatically*" and that "*I was not treated with empathy*". In terms of the experience of learning about fertility and preconception health, 114 (41.5%) comments showed appreciation for increased knowledge and awareness, but 30 (10.9%) stated that the content was superficial or needed more details. One respondent noted that "*It was informative and helpful. Although it would be sufficient for prior learning … it would be better for those who are trying to get pregnant to visit doctors for further information*".

Discussion

In our study, users who learned through conversation with an educational chatbot increased their fertility knowledge by 9 points (+15%) on the CFKS-J and had greater intentions to optimise their preconception health behaviours. Although improvement of fertility knowledge was smaller in the chatbot group (IG) than in the educational booklet group (CG1), the effects on behaviour modification were equivalent between the two groups. Currently, fertility awareness depends on different types of interventions – for example, from public health interventions delivered to many people to personalised one-to-one counselling delivered to fewer (Hvidman et al., 2014; Stern et al., 2013). In Japan, there are fertility awareness campaigns targeting young people (e.g., newlywed couples or those attending coming-of-age ceremonies) as well as clinics providing preconception care. Consultation fees at these clinics are not covered by public health insurance, however, and thus they involve extra expense for the people who use them. Our results suggest that new digital technology can provide more options for fertility and preconception health education delivered at the population level at a low cost. To improve knowledge of fertility health among people of reproductive age, further technical development to enable smooth and flexible communication is required.

The level of fertility knowledge improved considerably immediately after exposure to fertility information in the IG and CG1 (Figure 3A). These results align with previous findings that fertility knowledge consistently improves immediately after provision of information, irrespective of educational strategy, such as web-based documents (Wojcieszek and Thompson, 2013; Daniluk and Koert, 2015; Boivin et al., 2018a), video (Conceição et al., 2017), and face-to-face encounters (Garcia et al., 2016; Stern et al., 2013). Participants in the IG showed a 15% increase in fertility knowledge from the pre-test to the post-test, compared with a 24% increase in the CG1, in which participants received an in-depth booklet about female preconception health. One explanation could be that some participants in the IG did not experience enough conversation with the chatbot due to technical problems: the mean post-test scores of the 25 participants who reported insufficient exposure was 55 points, which was significantly lower than the post-test scores of the rest of the IG (i.e., 70 points; data not shown). Technical improvements to stabilize the chatbot system might further increase these knowledge gains.

Another explanation for the smaller-than-expected knowledge improvement in the IG could be that the predetermined communication did not meet the needs of the participants, thus they could not increase their knowledge. Although the scripted chatbot used natural language processing to understand users' responses, whenever the conversation veered from the predetermined scenarios, it responded, "*I'm sorry, I don't understand your question*". We could not identify each user's transcripts or the timing of drop-out, but we speculate that some participants in the IG might have given up on learning because they needed to follow all the chatbot instructions and answer questions. In contrast, those in CG1 could have skipped paragraphs in the PDF that contained information that they already knew, and focused only on what they wanted and needed to learn. Instead of scripted chatbots, artificially intelligent chatbots (Wall, 2018) could be built using big datasets (e.g., transcripts of patient-professional conversations), which may provide more appropriate and tailored information to users.

Significantly lower state anxiety in the IG suggests suitability of the chatbot for fertility awareness. We previously showed that provision of fertility information offers benefits of increased knowledge but also induces anxiety (Maeda et al., 2016). We replicated these results in the post-test anxiety scores of those in CG1 (Figure 3B). Yet, state anxiety in the IG was low, despite the knowledge increase. The seemingly non-relevant information given to CG2 (i.e., national pension system) also might have provoked anxiety because the declining birth rate in Japan could be a future threat to supporting an aging population (Nomura and Koizumi, 2016). Still, the post-test anxiety level of the IG was similar and even lower than that of control groups from a previous study (Maeda et al., 2016). One reason for low state anxiety in the IG could be attributed to a smaller educational effect (+15% versus +24% in fertility knowledge); participants in the IG did not learn enough to become stressed. However, the use of the chatbot itself might have alleviated the psychological stress, as confirmed by the feedback describing the chatbot as an easy, convenient, and casual tool that avoids embarrassment. Although further psychological evaluation of specific conversation that could make people feel anxious (e.g., for women, age they should start trying to conceive) is needed to determine if the chatbot achieved equivalent or larger educational effects than other methods, educational interventions that do not provoke anxiety can benefit users.

Online short education improved participants' intentions to participate in a wide range of preconception behaviours (Table 2). Substantial literature shows that preconception education and counselling improves maternal knowledge and behaviours, although effects on pregnancy outcomes remain unclear (Hussein et al., 2016;; Barker et al., 2018). In Japan, some well-known facts include the adverse effects of smoking, benefits of cervical check-ups, and the necessity of contraception, as shown in the relatively high proportions of participants who exhibited those behaviours compared with national statistics (Table 1). National statistics present similar data: the smoking rate among women in their 20s and 30s is 6%−9% (Ministry of Health, Labour, and Welfare, 2017b); the rate of biennial cervical check-ups is 42% (National Cancer Center of Japan); and contraception rates among married and unmarried women are 46% and 87%, respectively (National Institute of Population and Social Security Research, 2015). In this study, we also confirmed that Japanese participants' knowledge of "unfamiliar" preventive behaviours improved. Japan has low use rates of preconception folic acid (Ishikawa et al., 2018) and oral contraceptive pills (Yoshida et al., 2016). Another prominent concern is the extremely low rate of HPV vaccination (less than 1% among teenagers), likely because of a political change in 2013 that led to the suspension of proactive recommendations for the vaccine following intensive and sensational media coverage of unconfirmed adverse events (Hanley et al., 2015). In light of these health care challenges, our chatbot increased the percentage of participants who stated that they were "ready to take action" regarding their intake of folic acid, use of oral contraceptives, HPV vaccination status, and choice of primary ob/gyn doctor. Indeed, we need to assess actual behavioural changes through a follow-up study because knowledge is necessary but not always sufficient to change behaviour; for example, people need to feel susceptible to problems before they seek help (Fulford et al., 2013). However, a chatbot could at least be a useful strategy for promoting good health and preventing misunderstanding of health-related information in existing materials and for addressing misconceptions arising from mental models of the robustness of pregnancy (Fulford et al. 2014).

Recently, chatbots have been used in health care, such as teen health education (Crutzen et al., 2011), sexually transmitted infection screening (Kobori et al., 2018), nurse training (Shorey et al., 2019), chronic patient monitoring (Piau et al., 2019), genetic counselling (Schmidlen et al., 2019), and post-examination care (Goldenthal et al., 2019). Most of these one-armed studies assessed feasibility and reported positive feedback from users. Similar technology using virtual characters has shown promising results. For example, researchers at Boston University developed a virtual patient advocate named Gabby, who provides preconception health information and education using verbal and nonverbal communication. Users can respond to

Gabby by selecting and clicking on the button best representing their own responses (Gardiner et al., 2013). A six-month, randomized, controlled trial showed that preconception risk was lower among the Gabby user group, compared with a control group that received a letter listing personal health risks (Jack et al., 2015). In our randomized, controlled trial, a chatbot was designed to promote proactive learning through free-text input instead of selecting options. The results unexpectedly showed that although the chatbot produced significant knowledge gains, these gains were inferior to those produced by well-written material on its own. Currently, chatbot use in health education is limited, with diverse product specifications and study designs (e.g., two-armed randomised controlled trials, cross-overs). Results from these studies will help clarify the most effective specifications for using the technology (e.g., visual or auditory, concise or detailed, passive or active).

Another important implication is the need to investigate personal affinity or preference toward the use of technology. We found no moderation effects of sociodemographic factors (e.g., age, university education) or current pregnancy intention on knowledge increase of the intervention (data not shown). The lack of humanity and empathy perceived by some users also requires further engineering innovation. Notably, some participants reported that they preferred the chatbot as a convenient and easy way to talk about sensitive topics, which accords with previous studies suggesting that a virtual agent can alleviate negative feelings (Lucas et al., 2014; Palanica et al., 2019). On the other hand, some users evaluated the chatbot as lacking humanity or empathy. To improve this user experience and optimise the technology used in educational settings, further testing should include Think Aloud protocols or cognitive interviewing while interviewees are using the chatbot to access more deeply their thoughts about using the technology.

This study has some limitations. First, the use of social research panels could have caused selection bias associated with higher education (Haagen et al., 2003; Takahashi et al., 2011). In this study, to encourage participants to take enough time to learn, we told participants beforehand that the survey would include a one-hour learning session, which could have led to volunteer bias toward people who were more educated and more interested in childbearing. In fact, the mean pre-test knowledge score of the present participants was 61 points, which is equivalent to scores found by international studies (Bunting et al., 2013; Boivin et al., 2018a) but much higher than the average of 50 points found in Japan (Maeda et al., 2016). Also, the prevalence of participants who reported taking folic acid and oral contraceptives was higher in our study than in previous national data (Ishikawa et al., 2018; Yoshida et al., 2016). Second, although participants in all groups were instructed to close the study website before proceeding to a post-test survey, keeping the study material open was possible. The fact that participants in CG1 could have more easily looked for the post-test answers than those in the IG might have led to measurement bias. Third, the outcomes measured in this study were knowledge, intention, and psychological change immediately after exposure using mostly self-reported measurements. Long-term evaluations of hard outcomes (i.e., timing of first birth, actual behavioural change, and health of mothers and children), possibly accompanied with biomarkers for behaviour (e.g., cotinine measurement for smoking status), could be explored as benefits on these have been reported (Maeda et al., 2018). Fourth, due to the costs of development, we could not create an educational chatbot for men or all the people in this study. It is clear that information needs to be delivered to men as well. Finally, this study was conducted in Japan using a social research panel, and thus responses to fertility information and acceptance of digital technology might differ from those in other groups or cultures. Cultural relevance to educational strategies and new technologies could be an area of future research.

In conclusion, women who used an educational chatbot to learn about fertility and preconception health significantly increased their fertility knowledge and modified their intentions to optimise their preconception health immediately after exposure. However, the improvement in fertility knowledge was smaller than that of participants who read a well-written booklet, possibly because our chatbot had been in an early phase of development or because the evaluation included methodological limitations such as selection bias. Nevertheless, the impact on intentions and the finding that the chatbot did not provoke anxiety makes it a promising educational strategy for application at the population level. Further technical development and studies exploring personal affinity for technology in fertility awareness should be continued.

Acknowledgements

This study was funded by Pfizer Health Research Foundation (Grant Number 29-25) and joint research funding from the *1 More Baby Ohendan* foundation in Japan.

We thank the people involved in this study for their cooperation: board members of the incorporated foundation *1 More Baby Ohendan*, members of the *umuumu* project (Dentsu, Inc.), Mr. Yusuke Goto (Azest, Inc.), Ms. Megumi Kinsho, and Ms. Junko Hirayama (Akita University School of Medicine). This study was presented at the 36th Annual Meeting of the European Society of Human Reproduction and Embryology, July 5-8, 2020.

Authors' Roles

EM contributed to the conception and design of the study; participated in the acquisition, analyses, and interpretations of data; drafted all versions of the article; and approved the final version for publication. AM contributed to the conception and design of the study, the analysis and interpretation of data, all revisions, and the final approval for publication. JB and KN contributed to the analysis and interpretation of data, all revisions, and the final approval for publication. YK, HShirasawa, HSaito, and YT contributed to the conception and design of the study, all revisions, and the final approval for publication.

Conflict of Interest

EM reports joint research funding from a public interest, the incorporated foundation *1 More Baby Ohendan.*

References

Barker M, Dombrowski SU, Colbourn T, Fall CH, Kriznik NM, Lawrence WT et al. Intervention strategies to improve nutrition and health behaviours before conception. *Lancet* 2018; 391, 1853-1864

Boivin J, Koert E, Harris T, O'Shea L, Perryman A, Parker K, Harrison C. An experimental evaluation of the benefits and costs of providing fertility information to adolescents and emerging adults. *Hum Reprod* 2018a; 33 (7): 1247–1253.

Boivin J, Sandhu A, Brian K, Harrison C. Fertility-related knowledge and perceptions of fertility education among adolescents and emerging adults: a qualitative study. *Hum Fertil* 2018b, 1–9.

Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. *Fertil Steril* 2017;107(4):840-847.

Bunting L, Boivin J. Development and preliminary validation of the fertility status awareness tool: FertiSTAT. *Hum Reprod* 2010; 25(7): 1722–1733.

Bunting L, Tsibulsky I, Boivin J. Fertility knowledge and beliefs about fertility treatment: findings from the International Fertility Decision-making Study. *Hum Reprod* 2013; 28: 385–397.

Cardiff Fertility Studies Research Group. A Guide to Fertility. 2016. https://www.cardiff.ac.uk/data/assets/pdf_file/0008/1319570/Cardiff_University_A_Guide_to_Fertility_6-4-18.pdf (Accessed on 29 December 2019).

Conceição C, Pedro J, Martins MV. Effectiveness of a video intervention on fertility knowledge among university students: a randomised pre-test/post-test study. *Eur J Contracept Reprod Health Care* 2017; 22(2): 107–113.

Crutzen R, Peters GJ, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *J Adolesc Health* 2011; 48(5): 514–519.

Daniluk JC, Koert E. The other side of the fertility coin: a comparison of childless men's and women's knowledge of fertility and assisted reproductive technology. *Fertil Steril* 2013; 99: 839–846.

Daniluk JC, Koert E. Fertility awareness online: the efficacy of a fertility education website in increasing knowledge and changing fertility beliefs. *Hum Reprod* 2015; 30(2): 353–363.

Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal* 2015; 15(1): 292–300.

Fulford B, Bunting L, Tsibulsky I, Boivin J. The role of knowledge and perceived susceptibility in intentions to optimize fertility: findings from the International Fertility Decision-Making Study (IFDMS). *Hum Reprod*. 2013;28(12):3253-3262.

Fulford B, Macklon N, Boivin J. Mental models of pregnancy may explain low adherence to folic acid supplementation guidelines: a cross-sectional international survey. *Eur J Obstet Gynecol Reprod Biol*. 2014;176:99–103.

Garcia D, Vassena R, Prat A, Vernaeve V. Increasing fertility knowledge and awareness by tailored education: a randomized controlled trial. *Reprod Biomed Online* 2016; 32(1): 113–120.

Gardiner P, Hempstead MB, Ring L, Bickmore T, Yinusa-Nyahkoon L, Tran H et al. Reaching women through health information technology: the Gabby preconception care system. Am J Health Promot. 2013 Jan-Feb;27(3 Suppl):eS11-20.

Goldenthal SB, Portney D, Steppe E, Ghani K, Ellimoottil C. Assessing the feasibility of a chatbot after ureteroscopy. *M health* 2019; 5: 8.

Haagen EC, Tuil W, Hendriks J, de Bruijn RP, Braat DD, Kremer JA. Current Internet use and preferences of IVF and ICSI patients. *Hum Reprod* 2003;18: 2073–2078.

Habbema JDF, Eijkemans MJ, Leridon H, te Velde ER. Realizing a desired family size: when should couples start? *Hum Reprod* 2015; 30: 2215–2221.

Hammarberg K, Norman RJ, Robertson S, McLachlan R, Michelmore J, Johnson L. Development of a health promotion programme to improve awareness of factors that affect fertility, and evaluation of its reach in the first 5 years. *Reprod Biomed Soc Online* 2013; 4: 33–40.

Hanley SJ, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. *Lancet* 2015; 385, 2571.

Harper J, Hepburn J, Vautier G, Callander E, Glasgow T, Balen A, Boivin, J. Feasibility and acceptability of theatrical and visual art to deliver fertility education to young adults. *Hum Fertil (Camb)* 2019; 1–7. doi:10.1080/14647273.2019.1570354

Hussein N, Kai J, Qureshi N. The effects of preconception interventions on improving reproductive health and pregnancy outcomes in primary care: A systematic review. *Eur J Gen Pract* 2016; 22(1): 42–52.

Hvidman HW, Petersen KB, Larsen EC, Macklon KT, Pinborg A, Nyboe Andersen A. Individual fertility assessment and pro-fertility counselling; should this be offered to women and men of reproductive age? *Hum Reprod* 2014; 30(1): 9–15.

Ishikawa T, Obara T, Nishigori H, Nishigori T, Metoki H, Ishikuro M et al. Update on the prevalence and determinants of folic acid use in Japan evaluated with 91,538 pregnant women: the Japan Environment and Children’s Study. *J Matern Fetal Neonatal Med* 2018; 33(3): 427–436.

Jack BW, Atrash H, Coonrod DV, Moos MK, O'Donnell J, Johnson K. The clinical content of preconception care: an overview and preparation of this supplement. *Am J Obstet Gynecol* 2008; 199(6 Suppl 2): S266–S279.

Jack B, Bickmore T, Hempstead M, Yinusa-Nyahkoon L, Sadikova E, Mitchell S, et al. Reducing preconception risks among African American women with conversational agent technology. *J Am Board Fam Med* 2015; 28(4), 441-451.

Japan Society of Obstetrics and Gynecology, 2018. Dictionary for women and men, 2nd edition (in Japanese). http://www.jsog.or.jp/public/human_plus_dictionary/book_vol2.pdf (Accessed on 29 December 2019).

Kobori Y, Osaka A, Soh S, Okada H. Mp15-03 novel application for sexual transmitted infection screening with an AI chatbot. *J Urol* 2018; 199(4S): e189–e190.

Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*. 2010;31(5):702-755.

Lucas G, Gratch J, King A, Morency L. It’s only a computer: virtual humans increase willingness to disclose. *Comp Hum Behav* 2014; 37: 94–100.

Maeda E, Boivin J, Toyokawa S, Murata K, Saito H. Two-year follow-up of a randomized controlled trial: knowledge and reproductive outcome after online fertility education. *Hum Reprod* 2018; 33(11): 2035–2042.

Maeda E, Nakamura F, Kobayashi Y, Boivin J, Sugimori H, Murata K, Saito H. Effects of fertility education on knowledge, desires and anxiety among the reproductive-aged population: findings from a randomized controlled trial. *Hum Reprod* 2016; 31(9): 2051–2060.

Maeda E, Sugimori H, Nakamura F, Kobayashi Y, Green J, Suka M et al. A cross sectional study on fertility knowledge in Japan, measured with the Japanese version of Cardiff Fertility Knowledge Scale (CFKS-J). *Reprod Health* 2015; 12(1).

Malnory ME, Johnson TS. The reproductive life plan as a strategy to decrease poor birth outcomes. *J Obstet Gynecol Neonatal Nurs* 2011; 40(1): 109–121.

Mills M, Rindfuss RR, McDonald P, Velde E. Why do people postpone parenthood? Reasons and social policy incentives. *Hum Reprod* 2011; 17(6): 848–860.

Ministry of Education, Culture, Sports, Science and Technology. School Basic Survey 2008. https://www.e-stat.go.jp/dbview?sid=0003147040 (Accessed on 3 August 2020)

Ministry of Health, Labour, and Welfare. Nenkinseido no point (in Japanese) 2017a. https://www.mhlw.go.jp/topics/bukyoku/nenkin/nenkin/pdf/seido-h29-point.pdf. (Accessed on 29 December 2019).

Ministry of Health, Labour, and Welfare. National Health and Nutrition Survey. 2017b.

Nakazato K, Mizuguchi K. Development and validation of Japanese version of State–Trait Anxiety Inventory (STAI). *Jpn J Psychosom Med* 1982; 22: 107–112.

National Cancer Center. Cancer Registry and Statistics. https://ganjoho.jp/reg_stat/statistics/dl_screening/index.html (Accessed on 21 January 2020).

National Institute of Population and Social Security Research. The 15th National Fertility Survey. http://www.ipss.go.jp/site-ad/index_english/Survey-e.asp (Accessed on 8 August 2020).

Nomura K, Koizumi A. Strategy against aging society with declining birthrate in Japan. *Ind Health* 2016; 54(6), 477–479.

Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J Med Internet Res* 2019; 21(4): e12887.

Piau A, Crissey R, Brechemier D, Balardy L, Nourhashemi F. A smartphone Chatbot application to optimize monitoring of older patients with cancer. *Int J Med Inform* 2019; 128: 18–23.

Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983; 51(3), 390–395.

Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. *J Genet Couns* 2019;28(6):1166–1177.

Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. *J Med Internet Res* 2019; 21(10): e14658.

Spielberger CD, Gorsuch RL, Lushene RE. Manual for State-Trait Anxiety Inventory (Self-evaluating Questionnaire). Palo Alto. 1970.

Stein N, Brooks K. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. *JMIR Diabetes* 2017; 2(2): e28.

Stephenson J, Heslehurst N, Hall J, Schoenaker DA, Hutchinson J, Cade JE et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet* 2018; 391(10132): 1830–1841.

Stern J, Larsson M, Kristiansson P, Tydén T. Introducing reproductive life plan-based information in contraceptive counselling: an RCT. *Hum Reprod* 2013; 28(9): 2450–2461.

Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. *J Med Internet Res* 2011; 13: e110.

Thong EP, Codner E, Laven JSE, Teede H. Diabetes: a metabolic and reproductive disorder in women. *Lancet Diabetes Endocrinol*. 2020;8(2):134-149.

Trussell J, Aiken ARA, Micks E, Guthrie KA. Efficacy, safety, and personal considerations. In: Hatcher RA, Nelson AL, Trussell J, Cwiak C, Cason P, Policar MS, Edelman A, Aiken ARA, Marrazzo J, Kowal D, eds. Contraceptive technology. 21st ed. New York, NY: Ayer Company Publishers, Inc., 2018.

Twenge JM. iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy--and completely unprepared for adulthood--and what that means for the rest of us. Simon and Schuster, 2017.

van Dijk MR, Koster MPH, Willemsen SP, Huijgen NA, Laven JSE, Steegers-Theunissen RPM. Healthy preconception nutrition and lifestyle using personalized mobile health coaching is associated with enhanced pregnancy chance. *Reprod Biomed Online*. 2017;35(4):453-460.

Verbiest S, McClain E, Woodward S. Advancing preconception health in the United States: strategies for change. *Ups J Med Sci* 2016; 121(4): 222–226.

Wall M. Is this the year 'weaponised' AI bots do battle? BBC News. 5 January 2018 https://www.bbc.com/news/business-42559967 (Accessed on 29 December 2019).

Wojcieszek AM, Thompson R. Conceiving of change: a brief intervention increases young adults' knowledge of fertility and the effectiveness of in vitro fertilization. *Fertil Steril* 2013; 100: 523–529.

World Health Organization. Preconception care to reduce maternal and childhood mortality and morbidity. In Meeting Report and Packages of Interventions: WHO HQ, February 2012.

Yoshida H, Sakamoto H, Leslie A, Takahashi O, Tsuboi S, Kitamura K. Contraception in Japan: Current trends. *Contraception* 2016; 93(6): 475–477.

Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R et al. The international glossary on infertility and fertility care, 2017. *Hum Reprod* 2017; 32(9): 1786–1801.

Zelkowitz P, Miner S, O'Connell S, Robins S. An mHealth App Designed for Fertility Patients: From Conception to Pilot Testing. *Iproc* 2019;5(1):e15236.