Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Guideline-based conversational artificial intelligence versus a web patient guideline before breast cancer surgery: a randomized controlled trial

Chihiro Kuwahara
Okayama University Hospital
Masahide Koda
Okayama University
Yuko Takahashi
Okayama University Hospital
Takahiro Tsukioki
Okayama University Hospital
Shogo Nakamoto
Okayama University Hospital
Reina Maeda
Okayama University Hospital
Chika Sakudo
Okayama University Hospital
Marie Saji
Okayama University Hospital
Nozomi Saito
Okayama University Hospital
Yuki Fujiwara
Okayama University Hospital
Yuki Narahara
Okayama University Hospital
Maya Kosaka
Okayama University Hospital
Shutaro Kato
Okayama University Hospital
Asuka Mimata
Okayama University Hospital
Yusuke Otani
Okayama University Hospital
Maki Tanioka

Okayama University Hospital

Tadahiko Shien

tshien@md.okayama-u.ac.jp

Okayama University Hospital

Shinichi Toyooka

Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences

---

Article

Keywords: breast cancer, conversational artificial intelligence, chatbot, shared decision-making, patient education, randomized controlled trial

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11566242/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Conversational artificial intelligence may help patients access guideline information before breast cancer surgery. We evaluated whether a guideline-based chatbot improved patient-reported shared decision-making compared with a web patient guideline.

Methods

In this single-center randomized controlled trial, adults with clinical stage 0–II breast cancer scheduled for surgery before systemic therapy received access to a chatbot or the web version of the same patient guideline during outpatient waiting time. Allocation was stratified by age and diagnosis before referral. The primary outcome was the 9-item Shared Decision-Making Questionnaire score (0–100) after the surgical decision. The primary comparison used a Welch t test on available outcomes in the assigned groups. Post hoc analyses included adjustment for allocation strata, multiple imputation in all randomized participants, and sensitivity analyses for missing outcomes.

Results

All 92 randomized participants initiated their allocated resource (46 per group). Primary outcomes were available for 38 chatbot and 42 web-guideline participants. Mean scores were 87.61 and 89.86, respectively (chatbot minus web guideline, − 2.25 points; 95% confidence interval − 8.27 to 3.77; P = 0.46). Multiple imputation gave a difference of − 1.79 points (95% confidence interval − 7.86 to 4.28). No secondary outcome showed a statistically discernible difference. Detailed use measures were available for 18 participants per group. A documented service interruption affected free-text chatbot responses during recruitment; menu-based access remained available.

Conclusions

The chatbot as delivered did not demonstrate added benefit for patient-reported shared decision-making over a web version of the same patient guideline. Missing outcomes, incomplete use measurements, and variable conversational availability limit interpretation; equivalence and the effect of a consistently functioning interface were not established. Evaluations of conversational access alongside existing guideline resources should measure incremental patient benefit and outcomes matched to its intended clinical function.

Trial registration

Japan Registry of Clinical Trials, jRCT1062240109. Registered January 20, 2025, before enrollment began. https://jrct.mhlw.go.jp/latest-detail/jRCT1062240109

Background

A new diagnosis of breast cancer requires patients to understand unfamiliar information while making time-sensitive treatment decisions. For early-stage disease, the clinically appropriate options may differ in their consequences for local control, appearance, recovery, radiotherapy, and reconstruction. The importance that patients assign to these consequences can also differ. Shared decision-making (SDM) is intended to combine clinical evidence with each patient's preferences, circumstances, and desired role in the decision [1]. Patient decision aids make the decision explicit, describe options and their benefits and harms, and support consideration of personal values. Across health care settings, decision aids improve knowledge, risk perception, and participation in decision-making, although their effects vary by outcome and context [2]. For services already providing web-based patient guidelines, the practical question is whether conversational access adds benefit to shared decision-making.

Patient decision aids have been evaluated in breast cancer surgery. An early randomized trial showed that a decision aid improved knowledge and supported surgical treatment decision-making [3]. In the multicenter iCanDecide trial, an interactive, tailored website increased the probability of a high-quality locoregional treatment decision compared with a static website, primarily through improved knowledge rather than greater values concordance [4]. Evidence is not uniformly positive, however. A meta-analysis of web-based decision aids for postmastectomy breast reconstruction found possible improvements in knowledge, decisional conflict, and satisfaction, but not in informed choice, anxiety, or decision regret, with low or very low certainty for several outcomes [5]. A recent review of breast cancer decision aids similarly found modest benefits alongside substantial uncertainty [6]. In Japan, a field-tested decision aid for postmastectomy reconstruction demonstrated feasibility and acceptability, but randomized evidence in Japanese surgical pathways remains limited [7]. These findings suggest that the effect of digital support depends on the decision being addressed, the comparator, the intervention components, and the outcome selected.

Conversational artificial intelligence (AI) may extend digital decision support by allowing patients to ask questions in their own words, seek clarification, and revisit information outside the consultation. Retrieval-augmented generation can ground responses in a curated source, such as a patient guideline, rather than relying only on a model's general training data. The clinical evidence for breast cancer chatbots nevertheless remains narrow. In a randomized noninferiority trial using a fixed set of questions, chatbot answers were noninferior to physician-written answers for patient-rated satisfaction with information [8]. A prospective study described what patients asked a breast cancer chatbot over 1 year, without testing effects on SDM [9]. More recently, a guideline-grounded chatbot produced largely understandable and correct answers, but some responses were incomplete or potentially harmful [10]. These studies evaluated information provision and use, without testing effects on patient-reported shared decision-making during treatment.

Evidence for conversational agents in cancer care remains limited. A systematic review found promising effects across supportive care outcomes but also substantial heterogeneity, declining engagement, and a need for larger trials [11]. Reviews of conversational large language models in health care have identified similar concerns about safety, reliability, privacy, and clinical evaluation [12]. Evaluations should state how the digital function is expected to affect the selected patient outcome and should document the comparator, exposure, and implementation context [13].

Better access to guideline explanations could help patients prepare for consultation. Any effect on shared decision-making would also depend on how patients and clinicians discuss treatment options and preferences. The added value of conversational access therefore needs to be evaluated within the clinical pathway in which the information is used.

We hypothesized that access to a guideline-grounded conversational AI chatbot would improve patient-reported SDM compared with access to the web-based patient guideline. Secondary aims were to compare clinician-reported SDM, consultation time, patient satisfaction, clinician-rated understanding, postoperative decision regret, and subsequent inquiries.

Methods

Study design and setting

This single-center, parallel-group, randomized controlled trial was conducted in the Department of Breast and Endocrine Surgery at Okayama University Hospital, Japan. The trial was prospectively registered on January 20, 2025 (jRCT1062240109), before the first participant was enrolled on February 14, 2025. Protocol version 2.0 was dated August 6, 2025. Reporting follows CONSORT, CONSORT-EHEALTH, and CONSORT-AI guidance [14–16]. All methods were performed in accordance with the relevant guidelines and regulations.

Participants

Eligible patients were aged 18 years or older, had histologically confirmed clinical stage 0–II breast cancer, were scheduled for surgery before systemic therapy, could operate an iPad, and could provide written informed consent. Eligibility was based on the clinical information available at enrollment. Patients were ineligible if they could not use a tablet, were scheduled for neoadjuvant chemotherapy, or were considered unsuitable by an investigator. Researchers or physicians approached potentially eligible patients during outpatient visits, provided study information face-to-face, and obtained written informed consent.

Participant characteristics

Age, diagnosis before referral, clinical stage, hormone receptor status, human epidermal growth factor receptor 2 (HER2) status, breast surgery, and axillary surgery were summarized descriptively. Clinical stage and biomarker status reflected further diagnostic workup.

Randomization and blinding

Participants were allocated 1:1 using a computerized allocation system (Mujinwari), stratified by age (<65 or ≥65 years) and whether breast cancer had been diagnosed before referral. Permuted blocks of 4 or 6 were selected at random; the setting remained unchanged throughout the trial. A research team member registered participants and obtained their allocation from the system. The allocation was concealed until registration was complete. Participants knew which resource they used. Physicians and nurses assessing the consultation were not informed of allocation; participants were asked not to disclose their group and to close the intervention screen before consultation.

Information resources and delivery

Both groups received an iPad during outpatient waiting time, with at least 1 hour available and repeated access permitted. The chatbot group accessed an application on the miibo platform, using 81 knowledge-store entries drawn from the Japanese Breast Cancer Society's *Breast Cancer Clinical Practice Guideline for Patients, 2023 Edition*. The control group accessed the web version of the same patient guideline. Both resources used the same guideline source.

The chatbot offered free-text questions and menu-based access to explanations. It was intended to supplement the clinical consultation by helping patients obtain information. It did not autonomously diagnose disease, select treatment, or issue recommendations to clinicians. Treatment decisions remained with patients and their clinical teams. The intervention did not include structured values clarification or decision coaching. Shared decision-making was evaluated as a possible downstream effect of providing conversational information access.

Research staff explained how to operate the chatbot, provided consistent core information while allowing variation between staff members, and were available for technical questions. Participants were told that the resource did not replace physician consultation. They entered questions or selected menu items themselves; staff did not prescreen or modify inputs. No additional input-exclusion criteria, external reminders, or use caps were applied. The chatbot was restricted to trial participants and was not publicly accessible.

Technical availability and intervention-use measurements

Operational records document that free-text responses were unavailable from April 2 to April 22, 2026, after OpenAI credits were exhausted; menu-based access remained available. The response model was

temporarily changed to Claude Haiku on April 22 and restored to GPT-4o mini on April 24. These documented changes occurred during recruitment. Participant-level dates of resource use were unavailable, and platform records could not be linked to trial identifiers; the number of participants affected could not be determined. Protocol version 2.0 named OpenAI GPT and Google Gemini model families, but historical Gemini use could not be reconstructed.

The configuration inspected on July 28, 2026, used GPT-4o mini with vector retrieval from the guideline knowledge store. The system prompt instructed the model to answer from supplied guideline material and refer unanswered questions to staff or the treating physician. The separate knowledge-store-only response setting was not enabled. The dated configuration history was incomplete; Additional file 1, Table S3, gives the recovered settings and documented changes.

Intervention use was assessed from chat-turn counts and iPad screen time, with descriptive summaries for participants with recorded measurements. Platform operational logs, which contained no trial participant identifiers, were used to establish periods of service availability. Before the trial, the team checked responses against the patient guideline for accuracy and appropriateness. No formal trial-wide performance-error analysis or patient-linked assessment of harm attributable to chatbot responses was conducted.

Outcomes

The prespecified primary outcome was the continuous patient-reported 9-item Shared Decision-Making Questionnaire (SDM-Q-9) score after use of the allocated resource and the surgical treatment decision. Questionnaires were completed electronically on an iPad. The nine items were transformed to a 0–100 score, with higher values indicating greater perceived involvement in shared decision-making. The Japanese version has demonstrated reliability and validity [17].

Secondary outcomes were clinician-reported shared decision-making using the Japanese SDM-Q-Doc [18]; cumulative physician and nurse consultation time in minutes; patient satisfaction after iPad use; physician-rated patient understanding; unscheduled patient inquiries; and the Japanese Decision Regret Scale (DRS) [19–20]. The patient-satisfaction item was “Did you obtain the information you wanted by using the iPad?” The physician-rated understanding item was “Did the patient understand their condition and treatment options?” Both items used five response options: “strongly agree,” “agree,” “neither agree nor disagree,” “disagree,” and “strongly disagree” (translated from Japanese; score range, 1–5). Consultation time was recorded as the total across consultations for each participant; the number of consultations was summarized descriptively. DRS was assessed approximately 40 days after surgery, from 10 days before to 20 days after that target. Scores range from 0 to 100, with higher scores indicating greater regret. The continuous DRS score was the principal regret outcome. Post hoc descriptive categories were DRS 0, 1–25, and 26–100.

Sample size

The sample-size calculation assumed group means of 80.0 and 71.1, using SDM-Q-9 medians reported in a previous breast cancer trial as planning values [21]. Because that report did not provide standard deviations, values from a smaller breast reconstruction study [7] were multiplied by three, giving assumed standard deviations of 15.9 and 10.5. A two-sided alpha of 0.05 and 80% power required 76 participants (38 per group). The initial target was 80, allowing 5% attrition. After approximately 20% attrition was observed during recruitment, the target was amended to 100, with enrollment to stop once 80 participants who had not discontinued were available. No interim efficacy analysis was planned or performed.

Statistical analysis

The full analysis set comprised participants who initiated their assigned resource. The protocol specified Welch two-sample t tests and inclusion of participants who discontinued or had missing values, but did not specify an imputation procedure. The primary comparison used an available-case Welch t test on observed SDM-Q-9 scores in the original allocation groups, without imputation.

Continuous outcomes were summarized by means and standard deviations and by medians and interquartile ranges. Differences were calculated as chatbot minus web guideline, with 95% confidence intervals (CIs) and two-sided P values. Outcome-specific available cases were used for secondary comparisons. Unscheduled inquiries were recorded independently of primary-outcome completion. Secondary and post hoc analyses were exploratory, with no adjustment for multiple comparisons. For the post hoc proportion with DRS 0, an unadjusted risk difference and Wald 95% CI were calculated without a confirmatory hypothesis test.

Post hoc supportive analyses addressed allocation strata and missing primary outcomes. Linear regression adjusted for the two randomization strata. Multiple imputation by predictive mean matching used allocation group, age stratum, and diagnosis before referral to create 50 completed datasets including all randomized participants. Each completed dataset was analyzed using linear regression adjusted for the allocation strata, and estimates were combined using Rubin's rules. Departures from the missing-at-random assumption were examined by adding fixed offsets of −20, −10, −5, 0, 5, 10, and 20 points to imputed chatbot-group scores, bounded at 0 and 100. Additional file 1 presents these post hoc analyses. Analyses were performed in R version 4.5.2.

Use of language models

ChatGPT and Claude language models assisted with manuscript organization, English editing, and preparation and checking of the report. The authors are responsible for the study design, analyses, results, references, and interpretations.

Results

Participant flow and characteristics

Enrollment ended after 92 participants had been randomized, when 80 participants had completed the study without withdrawal, following the amended stopping rule. The final enrollment date was May 12, 2026; follow-up ended on July 9, 2026. Forty-six participants were assigned to each group, all initiated their assigned resource, and all remained in the full analysis set (Fig. 1).

Primary outcome measurements were available for 38 chatbot and 42 web-guideline participants. In the chatbot group, measurements were unavailable following a change to neoadjuvant therapy (n=4), difficulty using the iPad (n=2), surgery at another institution (n=1), or intervention withdrawal (n=1). In the web-guideline group, measurements were unavailable following neoadjuvant therapy (n=2), detection of bone metastasis after enrollment (n=1), or intervention withdrawal (n=1). These participants remained in the randomized cohort and in the multiple-imputation analysis.

Characteristics of the full analysis set are shown in Table 1. Mean ages were 53.98 years (SD 9.64) and 53.30 years (SD 10.43), respectively. Further diagnostic workup identified stage III disease in two chatbot participants and stage IV disease in one web-guideline participant. Eligibility had been assessed using the stage known at enrollment.

Primary outcome

Mean SDM-Q-9 scores were 87.61 (SD 12.95; n=38) in the chatbot group and 89.86 (SD 14.11; n=42) in the web-guideline group. The available-case difference was −2.25 points (95% CI −8.27 to 3.77; P=0.46; Table 2). Median scores were 91 (IQR 80–98) and 96 (IQR 84–100). Nine of 38 chatbot participants (23.7%) and 13 of 42 web-guideline participants (31.0%) scored 100 (Fig. 2).

Adjustment for randomization strata gave a difference of −2.06 points (95% CI −8.09 to 3.98; P=0.50). Multiple imputation in all 92 participants gave a difference of −1.79 points (95% CI −7.86 to 4.28; P=0.56). Across the examined offsets, differences ranged from −5.18 points (95% CI −11.63 to 1.28) at −20 to −0.11 points (95% CI −5.93 to 5.71) at +20. Additional file 1, Tables S1 and S2, reports these analyses.

Secondary outcomes

Table 2 gives the secondary-outcome comparisons. The SDM-Q-Doc difference was −2.38 points (95% CI −7.31 to 2.56; P=0.34). Differences in cumulative physician and nurse consultation time were −4.53 minutes (95% CI −11.26 to 2.21; P=0.18) and −2.80 minutes (95% CI −10.34 to 4.73; P=0.46), respectively. The differences in patient satisfaction, physician-rated understanding, and unscheduled

inquiries were −0.20 (95% CI −0.50 to 0.11), 0.22 (95% CI −0.07 to 0.51), and −0.09 (95% CI −0.22 to 0.04), respectively.

DRS data were available for 38 and 42 participants. Mean scores were 10.39 (SD 15.08) and 12.14 (SD 13.26), respectively, with a difference of −1.75 points (95% CI −8.10 to 4.61; P=0.59). Median scores were 0 (IQR 0–20) and 10 (IQR 0–25).

Available intervention-use measurements

Detailed use measurements were available for 18 of 46 participants (39.1%) in each group. Among those with measurements, the median chatbot turn count was 14.5 (IQR 6.25–23.75; range 2–43) and median screen time was 74 minutes (IQR 48.75–85.75; range 12–104). Median web-guideline screen time was 71.5 minutes (IQR 40–102.75; range 22–178).

Post hoc decision-regret categories

DRS 0 was reported by 21 of 38 chatbot participants (55.3%) and 17 of 42 web-guideline participants (40.5%), an absolute difference of 14.8 percentage points (95% CI −6.9 to 36.5). Scores of 1–25 occurred in 14 (36.8%) and 18 (42.9%) participants; scores of 26–100 occurred in three (7.9%) and seven (16.7%).

Discussion

Principal findings

This randomized comparison did not demonstrate better patient-reported shared decision-making with the chatbot as delivered than with a web patient guideline. Estimates from the available-case comparison, adjustment for allocation strata, and multiple imputation were similar, but the CIs remained compatible with lower or higher SDM scores.

The active comparator provided access to the same guideline source, making the comparison relevant to services considering a conversational interface alongside an existing web resource. The trial assessed its added clinical value during surgical care, when patients and clinicians were making treatment decisions.

Interpretation within the clinical pathway

The chatbot provided access to guideline explanations before consultation and was intended to help patients prepare for discussion with their clinical team. SDM-Q-9 evaluates the broader interaction in which patients and clinicians discuss options and preferences [17]. Better information access could contribute to that interaction, but its effect on SDM also depends on how the information is used during consultation.

Comparison with previous research

Patient decision aids generally improve knowledge and participation, although effects vary by intervention and outcome [2]. In iCanDecide, an interactive, tailored breast cancer website improved high-quality decisions largely through knowledge [4]. The present chatbot provided conversational guideline explanations without a structured values-clarification component. Its effect on knowledge or preparation for consultation may differ from its effect on global SDM.

Previous breast cancer chatbot studies evaluated satisfaction with information, use patterns, understandability, completeness, or potential harm [8–10]. The present randomized comparison evaluated patient-reported SDM during surgical care, providing evidence about the clinical outcome of offering conversational information access. Clinical implementation requires evidence on both technical function and the patient outcome that the function is intended to change [13].

Clinical implications and secondary outcomes

For services already providing web-based patient guidelines, these findings do not support assuming that adding conversational access will improve SDM. Evaluations should compare the added patient benefit with the existing resource and measure outcomes close to the tool's function, including resolution of questions and preparation for consultation.

Continuous decision regret did not show a statistically discernible difference. Decision regret was assessed approximately 40 days after surgery, so the results may differ from longer-term assessments. The higher proportion with DRS 0 was a post hoc finding with a wide CI and cannot establish a regret benefit.

The physician-consultation estimate was compatible with a reduction in cumulative time and with little or no difference. Consultation counts varied, and the trial did not establish a per-consultation efficiency effect. Future studies should assess clinician time together with comprehension, unresolved questions, subsequent contacts, and patient safety.

Strengths and limitations

Strengths include prospective registration, randomized allocation, validated Japanese outcome measures, and assessment within a clinical surgical pathway. Using an active comparator drawn from the same guideline source addresses the added value of a conversational interface where guideline information is already available.

Primary-outcome data were missing in 12 of 92 participants, with more missing outcomes in the chatbot group. Several measurements were unavailable after clinical events that changed the treatment pathway. The available-case primary comparison does not include observed outcomes from every randomized participant and may be biased if missingness is related to the outcome. Multiple imputation and the offset analyses depend on assumptions about missing outcomes, including those unavailable after a

change in treatment pathway. The imputation model used few predictors, and the offset analysis varied unobserved chatbot scores while retaining the imputation assumptions for missing web-guideline scores.

All randomized participants initiated their assigned resource, but detailed chat counts and screen times were available for only 18 participants per group. These measurements cannot establish engagement across the full cohort or distinguish limited use from incomplete measurement [22]. Free-text responses were interrupted during recruitment, the response model changed temporarily, and menu-based access continued. Platform records could not be linked to participants, so the number affected was unknown. Clinical effectiveness cannot be separated from implementation fidelity, and the effect of a consistently available conversational interface remains uncertain. Future trials should prospectively record use and technical changes alongside clinical outcomes [23].

Scores were concentrated near the upper end in both groups, which may have reduced sensitivity to an additional improvement. Knowledge of personally relevant options, resolution of questions, preparedness for consultation, and baseline decisional conflict were not directly measured. Baseline decisional need, perceived choice, health literacy, and preferred decision role were also not measured. Effects on these outcomes and explanations based on patient characteristics or cultural differences remain untested. Trial-wide response errors and related harms were not formally assessed.

The trial was not powered for secondary or subgroup effects. Participants were unblinded, and clinicians may have inferred allocation. The single-center sample limits precision and generalizability. Tablets, instructions, and technical support were provided by research staff; routine implementation may offer different resources and support.

Conclusions

Under the delivery conditions studied, offering conversational access to a patient guideline did not demonstrate added benefit for patient-reported SDM over its web version before breast cancer surgery. Missing outcomes, incomplete use measurements, and variable conversational availability limit interpretation, and equivalence was not established. Decisions about adopting conversational interfaces alongside existing digital guidelines require evidence of incremental patient benefit, with reliable delivery and outcomes matched to the intended clinical function.

Abbreviations

AI: artificial intelligence; HER2: human epidermal growth factor receptor 2; CI: confidence interval; DRS: Decision Regret Scale; IQR: interquartile range; SD: standard deviation; SDM: shared decision-making; SDM-Q-9: 9-item Shared Decision-Making Questionnaire; SDM-Q-Doc: physician version of the Shared Decision-Making Questionnaire.

Declarations

Ethics approval and consent to participate

The ethics committee of Okayama University approved the study (rin2412-001). All participants provided written informed consent. Study data used participant identification codes, with directly identifying information removed from the analysis data and the linkage file stored separately with password protection. Participants received no payment and incurred no additional research-related costs.

Consent for publication

Not applicable.

Availability of data and materials

The datasets are not publicly available because of privacy and ethical restrictions. Requests for access will be considered by the corresponding author and study team in accordance with the ethical approval and informed consent requirements. The study protocol is available from the corresponding author on reasonable request. Intervention specifications and supportive analyses are provided in Additional file 1. The miibo platform is proprietary; its source code was not available to the study team.

Competing interests

The authors declare no competing interests. The study team had no relationship with the developer or owner of miibo.

Funding

No financial support was received for this work.

Authors' contributions

CK, MK (Koda), YT, and TS conceived the study. CK and MK (Koda) drafted the initial manuscript. YT and TS critically revised the manuscript. CK, YT, TT, SN, RM, CS, MS, NS, YF, YN, MK (Kosaka), SK, AM, YO, MT, TS, and ST participated in conducting the randomized trial. TT, SN, RM, CS, MS, NS, YF, YN, MK (Kosaka), SK, AM, YO, MT, and ST reviewed the manuscript and provided supervision. MK (Koda) contributed to the statistical analysis and interpretation. MK (Koda) refers to Masahide Koda; MK (Kosaka) refers to Maya Kosaka.

Acknowledgements

The authors thank Toshiharu Mitsuhashi, Research Associate, Medical Development Field, Center for Innovative Clinical Medicine, Okayama University, for advice on the statistical analysis.

References

1. Elwyn, G. et al. Shared decision making: a model for clinical practice. *J. Gen. Intern. Med.* **27** (10), 1361–1367. 10.1007/s11606-012-2077-6 (2012).
2. Stacey, D. et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst. Rev.* **1** (1), CD001431. 10.1002/14651858.CD001431.pub6 (2024).
3. Whelan, T. et al. Effect of a decision aid on knowledge and treatment decision making for breast cancer surgery: a randomized trial. *JAMA* **292** (4), 435–441. 10.1001/jama.292.4.435 (2004).
4. Hawley, S. T. et al. Improving breast cancer surgical treatment decision making: the iCanDecide randomized clinical trial. *J. Clin. Oncol.* **36** (7), 659–666. 10.1200/JCO.2017.74.8442 (2018).
5. Yu, L. et al. Assessing the content and effect of web-based decision aids for postmastectomy breast reconstruction: systematic review and meta-analysis of randomized controlled trials. *J. Med. Internet Res.* **26**, e53872. 10.2196/53872 (2024).
6. Vasan Thakumar, A. et al. Effectiveness of decision aids on decisional conflict and treatment or disease knowledge among women diagnosed with breast cancer: a systematic review and meta-analysis. *J. Cancer Surviv Published online July*. **3** 10.1007/s11764-026-02075-5 (2026).
7. Sowa, Y. et al. Development and implementation of a decision aid for post-mastectomy breast reconstruction for Japanese women with breast cancer: a field-testing study. *Breast Cancer*. **30** (4), 570–576. 10.1007/s12282-023-01447-4 (2023).
8. Bibault, J. E. et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. *J. Med. Internet Res.* **21** (11), e15787. 10.2196/15787 (2019).
9. Chaix, B. et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer*. **5** (1), e12856. 10.2196/12856 (2019).
10. Boie, S. D. et al. Evaluating a chatbot as a companion for patients with breast cancer: collaborative pilot study. *JMIR Cancer*. **11**, e68426. 10.2196/68426 (2025).
11. Jiang, X. H., Yuan, X. H., Zhao, H. & Peng, J. S. The effectiveness and feasibility of conversational agents in supporting care for patients with cancer: systematic review and meta-analysis. *J. Med. Internet Res.* **27**, e76968. 10.2196/76968 (2025).
12. Wang, L. et al. Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. *J. Med. Internet Res.* **26**, e22769. 10.2196/22769 (2024).
13. Bolinger, E. & Tyl, B. Key considerations for designing clinical studies to evaluate digital health solutions. *J. Med. Internet Res.* **26**, e54518. 10.2196/54518 (2024).

14. Hopewell, S. et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ* **389**, e081123. 10.1136/bmj-2024-081123 (2025).
15. Eysenbach, G. & CONSORT-EHEALTH Group CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J. Med. Internet Res.* **13** (4), e126. 10.2196/jmir.1923 (2011).
16. Liu, X. et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ* **370**, m3164. 10.1136/bmj.m3164 (2020).
17. Goto, Y. et al. Psychometric evaluation of the Japanese 9-item Shared Decision-Making Questionnaire and its association with decision conflict and patient factors in Japanese primary care. *JMA J.* **3** (3), 208–215. 10.31662/jmaj.2019-0069 (2020).
18. Goto, Y. et al. Association between physicians' and patients' perspectives of shared decision making in primary care settings in Japan: the impact of environmental factors. *PLoS One*. **16** (2), e0246518. 10.1371/journal.pone.0246518 (2021).
19. Brehaut, J. C. et al. Validation of a Decision Regret Scale. *Med. Decis. Mak.* **23** (4), 281–292. 10.1177/0272989X03256005 (2003).
20. Tanno, K., Bito, S., Isobe, Y. & Takagi, Y. Validation of a Japanese version of the Decision Regret Scale. *J. Nurs. Meas.* **24** (1), E44–E54. 10.1891/1061-3749.24.1.E44 (2016).
21. Søndergaard, S. R. et al. Shared decision making with breast cancer patients: does it work? Results of the cluster-randomized, multicenter DBCG RT SDM trial. *Radiother Oncol.* **193**, 110115. 10.1016/j.radonc.2024.110115 (2024).
22. Perski, O., Blandford, A., West, R. & Michie, S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav. Med.* **7** (2), 254–267. 10.1007/s13142-016-0453-1 (2017).
23. Mohr, D. C. et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. *J. Med. Internet Res.* **17** (7), e166. 10.2196/jmir.4391 (2015).

Tables

Table 1. Characteristics of the randomized cohort

Characteristic	Category	Chatbot (n=46)	Web guideline (n=46)
Age, years	Mean (SD)	54.0 (9.6)	53.3 (10.4)
Age stratum, n (%)	<65 years	37 (80.4)	39 (84.8)
	≥65 years	9 (19.6)	7 (15.2)
Breast cancer diagnosed before referral, n (%)	No	18 (39.1)	17 (37.0)
	Yes	28 (60.9)	29 (63.0)
Clinical stage recorded after further workup, n (%)	0	13 (28.3)	7 (15.2)
	I	16 (34.8)	27 (58.7)
	II	15 (32.6)	11 (23.9)
	III	2 (4.3)	0 (0.0)
	IV	0 (0.0)	1 (2.2)
Hormone receptor status, n (%)	Negative	6 (13.0)	6 (13.0)
	Positive	39 (84.8)	40 (87.0)
	Missing	1 (2.2)	0 (0.0)
HER2 status, n (%)	Negative	31 (67.4)	38 (82.6)
	Positive	5 (10.9)	3 (6.5)
	Missing	10 (21.7)	5 (10.9)
Breast surgery, n (%)	Partial mastectomy	17 (37.0)	15 (32.6)
	Total mastectomy without reconstruction	13 (28.3)	17 (37.0)
	Total mastectomy with reconstruction	13 (28.3)	13 (28.3)
	Missing	3 (6.5)	1 (2.2)
Axillary surgery, n (%)	None	7 (15.2)	3 (6.5)
	Sentinel lymph node biopsy	31 (67.4)	36 (78.3)
	Axillary lymph node dissection	5 (10.9)	6 (13.0)
	Missing	3 (6.5)	1 (2.2)
Number of consultations, visits	Mean (SD)	1.7 (0.51)	1.8 (0.68)

Values are n (%) unless otherwise stated. Clinical stage and biomarker status reflect further diagnostic workup; clinical stage at enrollment determined eligibility. Age stratum and diagnosis before referral were randomization strata. SD, standard deviation.

Table 2. Primary and secondary outcomes by randomized group

Outcome	Chatbot	Web guideline	Difference, chatbot minus web (95% CI)	P value
SDM-Q-9	38; 87.61 (12.95); 91 (80-98)	42; 89.86 (14.11); 96 (84-100)	−2.25 (−8.27 to 3.77)	.46
SDM-Q-Doc	38; 86.55 (11.80); 88 (78-97.5)	42; 88.93 (10.16); 90 (80-99)	−2.38 (−7.31 to 2.56)	.34
Physician consultation, min	38; 25.71 (14.53); 23 (15-35.25)	42; 30.24 (15.72); 27.5 (18.25-41.5)	−4.53 (−11.26 to 2.21)	.18
Nurse consultation, min	38; 22.05 (16.06); 20 (13.5-23)	42; 24.86 (17.81); 20.5 (12.75-31.5)	−2.80 (−10.34 to 4.73)	.46
Patient satisfaction	38; 3.95 (0.77); 4 (4-4)	42; 4.14 (0.57); 4 (4-4)	−0.20 (−0.50 to 0.11)	.20
Physician-rated understanding	38; 4.32 (0.53); 4 (4-5)	42; 4.10 (0.76); 4 (4-5)	0.22 (−0.07 to 0.51)	.13
Decision Regret Scale	38; 10.39 (15.08); 0 (0-20)	42; 12.14 (13.26); 10 (0-25)	−1.75 (−8.10 to 4.61)	.59
Number of inquiries	41; 0.05 (0.22); 0 (0-0)	43; 0.14 (0.35); 0 (0-0)	−0.09 (−0.22 to 0.04)	.16

Values show n; mean (SD); median (IQR). Differences are chatbot minus web guideline, with 95% CIs, using available outcomes in the original allocation groups. SDM-Q-9 was the only primary outcome. Other comparisons were exploratory, without adjustment for multiplicity. Unscheduled inquiries were recorded independently of primary-outcome completion. Scales: SDM-Q-9, SDM-Q-Doc, and DRS, 0–100; satisfaction and physician-rated understanding, 1–5. Consultation times are cumulative per participant. CI, confidence interval; DRS, Decision Regret Scale; IQR, interquartile range; SD, standard deviation.

Figures

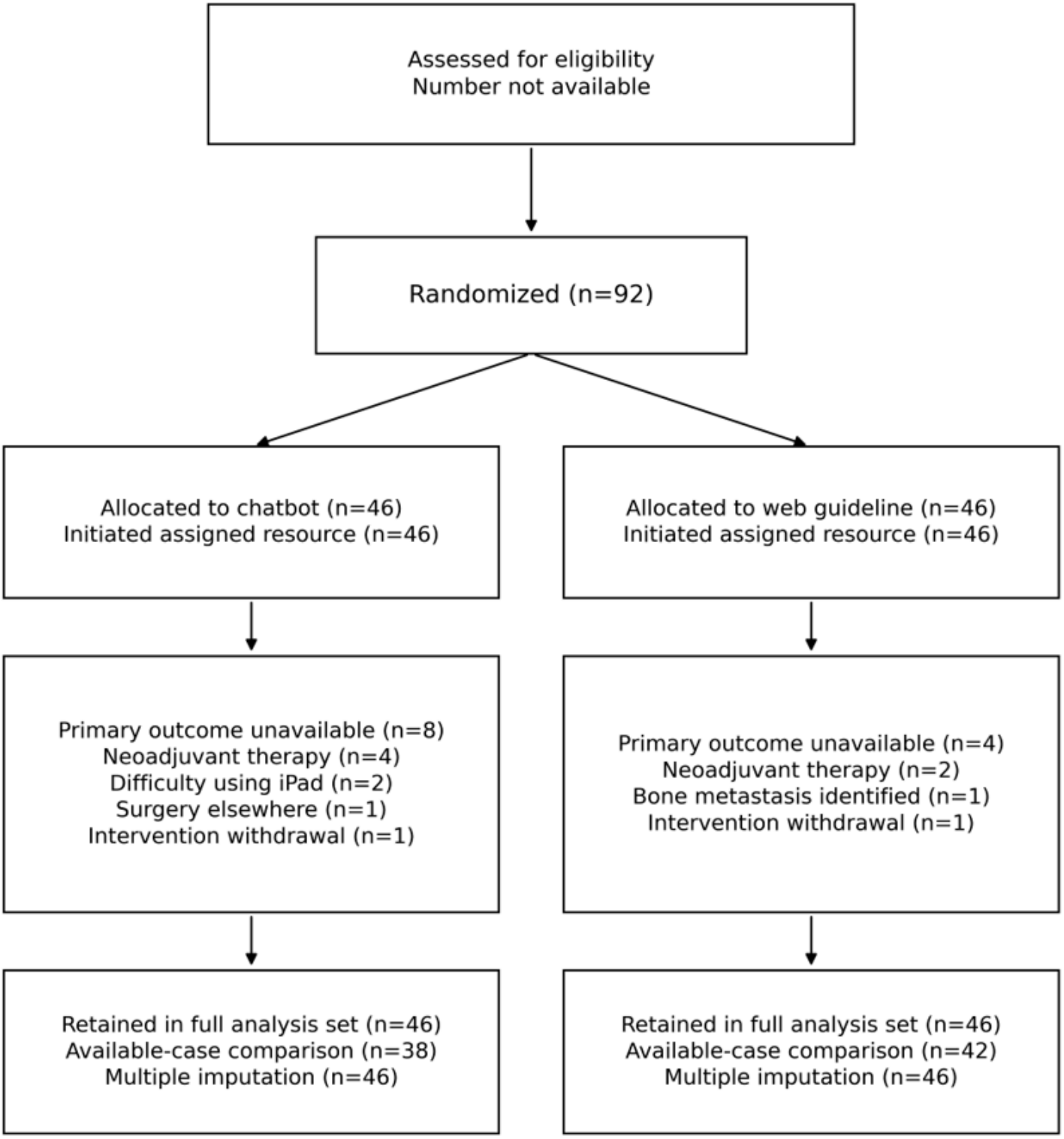


Figure 1

Participant flow and outcome availability. All 92 randomized participants initiated their assigned information resource and remained in the full analysis set. The available-case primary comparison included 38 chatbot and 42 web-guideline participants; the multiple-imputation analysis included all 46 participants per group. Reasons for missing primary outcomes are shown.

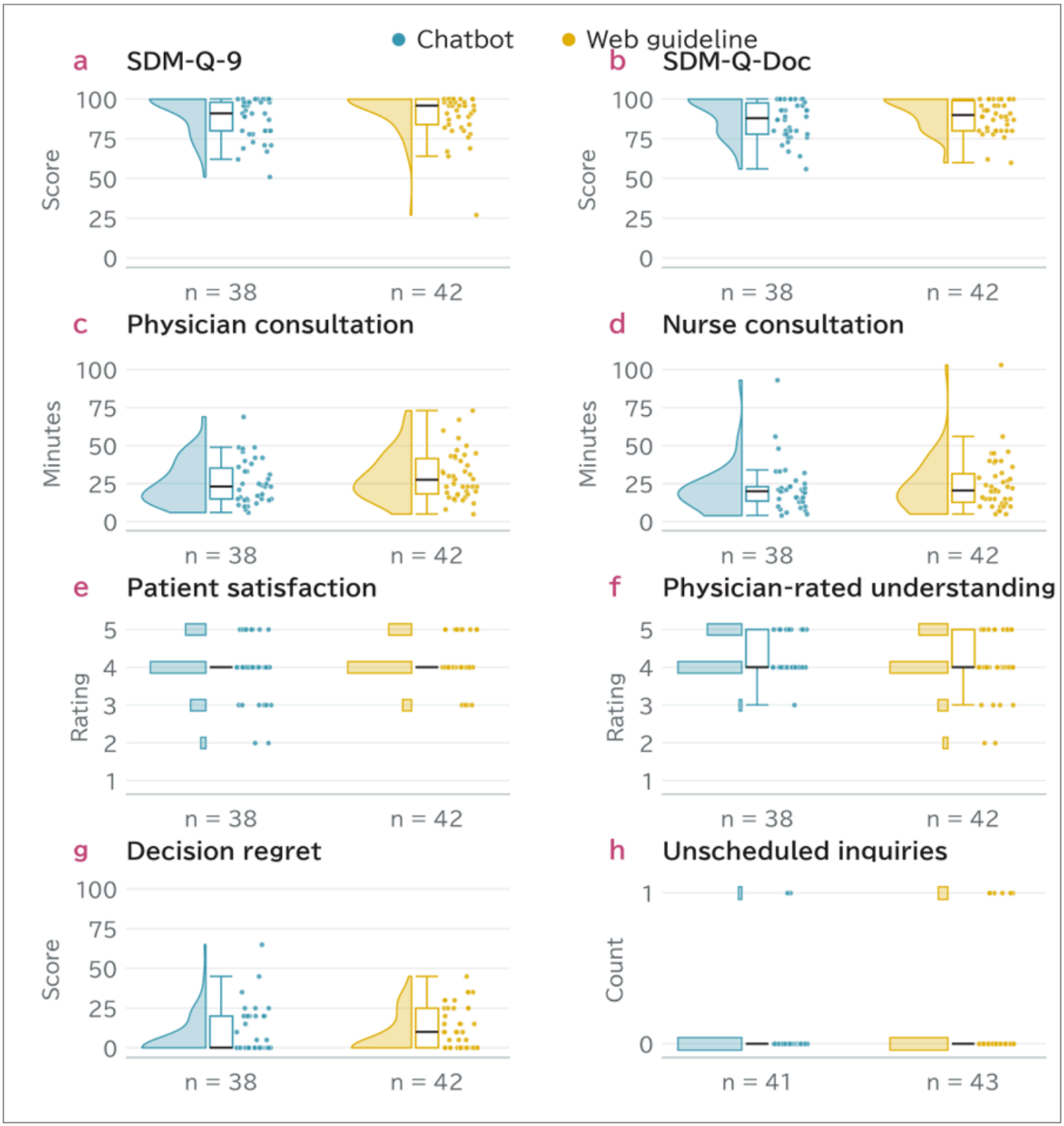


Figure 2

Distributions of primary and secondary outcomes. Half-violins show kernel density estimates for continuous outcomes; one-sided profiles show observed relative frequencies for discrete ratings and counts. Boxes show interquartile ranges and medians; whiskers extend to observations within 1.5 times the interquartile range. Points show individual participants, with horizontal jitter for visibility. Labels show outcome-specific sample sizes. Matched scales use the same axis limits. Consultation times are

cumulative per participant. DRS, Decision Regret Scale; SDM-Q-9, 9-item Shared Decision-Making Questionnaire; SDM-Q-Doc, physician version of the questionnaire.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- additionalfile1.docx