

Original Paper

Preliminary Evaluation of a Large Language Model–Powered Chatbot for Osteoporosis Self-Management Education: Formative Randomized Controlled Trial

Jinling Huang^{1,2}, MD; Xiaolian Xin¹, BSc; Chunyan He¹, BSc; Haoyan Xiong¹, MD; Wenjun Pang¹, BSc; Shaobo Pang¹, BSc; Wei Sun², MD; Xianghua Ding², PhD

¹Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

²University of Glasgow, Glasgow, Scotland, United Kingdom

Corresponding Author:

Haoyan Xiong, MD

Affiliated Hospital of Guangdong Medical University

No. 57, Renmin Avenue South, Xiashan District

Zhanjiang, 524000

China

Phone: 86 18316472969

Email: 810397996@qq.com

Abstract

Background: With the increasing burden of chronic diseases, self-management education (SME) is crucial. Traditional SME based on face-to-face delivery by clinicians is resource-intensive, and general digital tools such as web-based platforms often provide limited interactivity for patient learning. Although chatbots based on large language models (LLMs) show promise in interactivity, their real-world effectiveness lacks empirical evidence.

Objective: This study aimed to explore the feasibility and preliminary effectiveness of an LLM-based chatbot specifically designed for osteoporosis SME.

Methods: A formative randomized controlled trial was conducted in a tertiary hospital from February 2024 to March 2025. Adults aged ≥ 18 years with osteoporosis were recruited and randomly assigned (1:1) to either the intervention (OPBot) group or a control group receiving traditional health education. The chatbot provided interactive educational content and question-and-answer support, while the control group received face-to-face education and written materials. Osteoporosis knowledge was assessed using the Osteoporosis Knowledge Assessment Tool at baseline and discharge. Nurses' time spent on health education was self-recorded during each intervention session and aggregated across sessions. Adherence to disease management was assessed at 1, 3, and 6 months after discharge via telephone using Likert-scale questionnaires. The reliability of OPBot responses was evaluated by 2 clinician assessors using a 5-point Likert scale, with interrater agreement calculated using Cohen κ . Group comparisons were conducted using 2-tailed independent t tests, Mann-Whitney U tests, and chi-square tests, and adherence outcomes were analyzed using mixed-effects models.

Results: A total of 100 participants were randomized; 12% (12/100) were excluded due to loss to follow-up, refusal of the second knowledge assessment, or death, leaving 88% (88/100) participants for analysis ($n=45$, 51.1% in the OPBot group and $n=43$, 48.9% in the control group). The OPBot group showed significantly higher postintervention knowledge scores than the control group (median 80.0, IQR 70.0-89.0 vs median 75.0, IQR 65.5-80.0; $P=.01$). Nurses in the OPBot group spent lesser time on SME than those in the control group (median 5.0, IQR 2.0-17.0 vs median 23.0, IQR 20.0-25.0 minutes; $P<.001$). For adherence outcomes, a significant group \times time interaction was observed for calcium supplement intake (odds ratio 1.49, 95% CI 1.08-2.06; $P=.02$), indicating differing adherence trajectories over time. The OPBot group also showed higher odds of consuming calcium-rich foods across time points (odds ratio 2.87, 95% CI 1.04-7.89; nominal $P=.04$), although this association did not remain significant after Holm correction. No significant effects were observed for sun exposure ($P=.56$), exercise ($P=.79$), or total adherence scores ($P=.33$). In the question-and-answer module, most OPBot responses were rated as highly reliable 89.4% (76/85), with high interrater agreement (Cohen $\kappa=0.83$).

Conclusions: LLM-based chatbots specifically designed for osteoporosis SME may improve patient knowledge, supporting adherence behaviors, and reducing healthcare workload. However, further large-scale studies are needed to confirm these findings.

KEYWORDS

chatbots; large language models; self-management education; osteoporosis; adherence; patient education

Introduction

The global prevalence of chronic conditions is an increasing concern. In 2024, the World Health Organization reported that chronic conditions accounted for more than two-thirds of deaths worldwide [1]. In response to increasing pressure on health care services, self-management plays a crucial role in addressing these conditions, involving an individual's ability to manage symptoms, treatments, physical and psychosocial impacts, and necessary lifestyle changes [2]. Self-management education (SME) equips individuals with the essential knowledge and skills to manage their conditions effectively [3], making it an increasingly vital component of health care [4]. Without adequate education, individuals may struggle to understand their condition and lack management skills [5].

Traditionally, SME has been delivered by health care staff in medical settings such as hospitals or clinics, primarily through face-to-face lectures [6]. However, this approach is time-intensive and places a considerable burden on health care professionals. This burden is further exacerbated by ongoing workforce shortages and increasing service demands in modern health care systems [7]. In response to these limitations, online education, such as web-based platforms, mobile health apps, and structured e-learning programs, allows patients to learn at their convenience, in a more relaxed state, without being restricted by time or location [8]. However, these approaches often face challenges related to usability and limited interactivity, which may lead to low user engagement and reduced sustained use, thereby hindering their broader adoption [9].

In contrast, chatbots engage with users through natural and familiar conversations, enhancing their interactivity [10-12]. However, traditional chatbot systems, whether rule-based or based on conventional machine learning and natural language processing techniques, often rely on predefined scripts, limited training data, or shallow linguistic representations [13,14], making it challenging to effectively handle complex, open-ended, or personalized queries [13,15]. More recently, chatbots have increasingly been powered by large language models (LLMs), enabling more flexible, context-aware, and generative interactions [16,17], and have been explored across a range of scenarios, including mental health support [18,19] and patient education or counseling [20-22]. In the context of patient education in particular, prior studies examined the quality and understandability of responses of general LLM-powered chatbots and found that although they can provide general health-related information with reasonable quality and support patient understanding, they are limited for more specialized questions [23-26]. At the same time, current research has largely focused on evaluating system performance, including information quality and response accuracy [27,28], with limited empirical evidence on their effectiveness in practice [29], including self-management behavioral changes and workload

demands on health care professionals while maintaining reliable and clinically appropriate responses.

To address these gaps, we developed an LLM-based chatbot designed and specifically trained to deliver SME by providing interactive educational content and real-time question-and-answer support, focusing on osteoporosis. Osteoporosis is a skeletal condition characterized by low bone density and impaired bone metabolism, leading to fragile bones and an increased risk of fractures [30]. Globally, osteoporosis accounts for 70% of fragility fractures, which often result in chronic pain, reduced physical activity, and significant levels of disability and dependence [31,32]. Osteoporosis can be managed proactively to prevent fractures, and even after an initial fracture, interventions can reduce the likelihood of subsequent fractures [33]. The International Osteoporosis Foundation (IOF) highlights that preventive strategies often include lifestyle modifications. Without such management, the risk of a second fracture within 2 years is notably high. However, many patients lack awareness of the importance of lifestyle changes, often relying solely on medical treatments or struggling to effectively manage the disease [34,35].

Using osteoporosis as a case study, this research aimed to explore the feasibility and preliminary effectiveness of chatbot-supported SME for chronic disease management. We conducted a formative randomized controlled trial to (1) explore its potential for gaining self-management knowledge, (2) examine its possible influence on self-management adherence, and (3) assess its potential to reduce nurses' workload, while also evaluating the reliability of chatbot-generated responses.

Methods

Ethical Considerations

Prior to recruitment and data collection procedures, this study was approved by the institutional review board of the Affiliated Hospital of Guangdong Medical University (PJKT2024-216). All participants provided written informed consent before data collection and were informed that all study data would be deidentified prior to analysis. The data were stored on secure, password-protected online platforms (including BigQuery for OPBot usage data and the Wenjuanxing platform by Tencent for questionnaire data). Access to the data was restricted to authorized members of the research team to ensure participant privacy and data security, in accordance with institutional policies and applicable data protection regulations. Participants did not receive any financial compensation. However, all participants in both groups were provided with equivalent access to free follow-up services and osteoporosis-related health consultations during the study period as part of standard study support.

Study Design and Setting

A formative randomized controlled trial was conducted in a tertiary hospital from February 2024 to March 2025.

Participant Recruitment

The study recruited patients from February 2024 to September 2024, with follow-up completed in March 2025. Patients diagnosed with osteoporosis were recruited from the orthopedic ward of the study hospital. Eligible patients were approached by a research nurse after the diagnosis of osteoporosis had been confirmed and were invited to participate in the study during their hospitalization. Participants' personal information was obtained from the hospital's electronic medical record system. Recruitment took place in the hospital ward and continued until the target sample size was reached.

The inclusion criteria were (1) aged 18 years or older and (2) diagnosed with osteoporosis. The exclusion criteria were as follows: (1) unwilling or unable to use smartphones, (2) hearing impairment, (3) impaired consciousness, (4) critical illness, (5) paralyzed in bed, and (6) language barriers.

Sample Size Justification

The sample size for this formative randomized controlled trial was determined based on methodological guidelines for pilot and feasibility studies, rather than formal hypothesis testing. The target sample size for this study was 100 participants (50 per group), determined based on practical factors, such as available resources and recruitment capacity, to support the assessment of feasibility outcomes (including retention, compliance, and acceptability) and to provide an initial estimate for future confirmatory trials.

Randomization, Concealment, and Blinding

This study was designed as a randomized controlled trial. Researcher A assessed participants' eligibility, obtained written informed consent, and collected baseline data. Researcher B, who was not involved in recruitment, generated the random sequence using Microsoft Excel. To ensure allocation concealment, the sequence was placed in sequentially numbered, opaque, sealed envelopes. Prior to the start of the intervention, participants were randomly assigned in a 1:1 ratio to either the intervention group or the control group according to their enrollment order and the random code indicated on the envelope. Each group included 50 participants.

Due to the nature of the intervention, blinding of participants and the intervention coordinator responsible for delivering the intervention was not feasible, as only participants in the intervention group had access to OPBot. However, to minimize potential bias, the data analyst remained blinded to group allocation during statistical analysis.

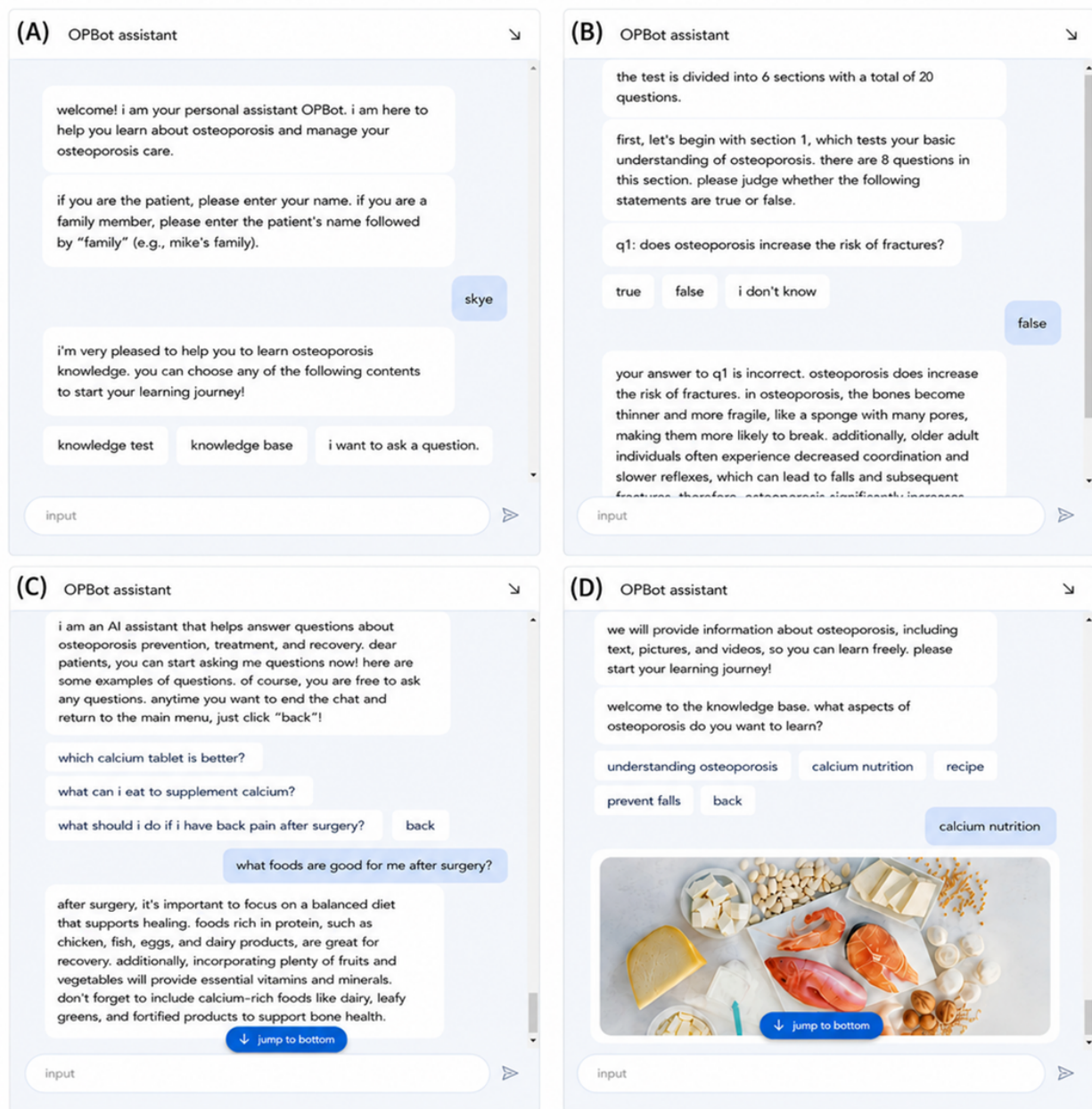
Intervention

Education Chatbot: OPBot

OPBot is a web-based application developed collaboratively by clinical staff and computer engineers, consisting of 3 modules (Figure 1). Among them, a knowledge base module was added to deliver osteoporosis knowledge and self-management resources. Furthermore, a knowledge assessment module was introduced. This module is designed to elicit learners' beliefs and theories, promoting the integration of new and refined concepts into their belief systems [35]. The knowledge assessment module provides educational feedback based on users' responses, enabling the recommendation of more tailored learning content and enhancing the effectiveness of health education. This also addresses the limitation in Chinese clinical practice where time constraints often hinder pre-education assessments. It uses content adapted from the Osteoporosis Knowledge Test [36] and translated into Chinese. The user-friendly design mirrors real-world clinical assessments.

Finally, the question-and-answer module is designed to provide an efficient and convenient solution to address patients' concerns during hospitalization and to offer ongoing support after discharge. OPBot's training materials were sourced from the IOF and clinical consultation records. The IOF content was reviewed, translated, and adapted by orthopedic nursing experts to align with the Chinese health care context. The clinical consultation records initially included 361 entries related to osteoporosis. After removing duplicates and irrelevant content, 128 (35.5%) refined entries were ultimately selected to train OPBot. Clinical evaluation by 2 nurses (a specialized nurse and a case manager) assessed the responses generated by the question-and-answer module to ensure practical relevance. The final suitability rate was 86.7%, indicating a moderately high level of suitability. However, given the need for precision in health care, OPBot should be used with ongoing involvement from health care providers, for example, by reviewing OPBot's responses daily and providing timely interventions when necessary. Details of the OPBot development protocol and evaluation process are provided in [Multimedia Appendix 1](#).

Figure 1. Overview of OPBot interface (translated into English for the paper). (A) Main interface, (B) knowledge assessment module, (C) question-and-answer module, and (D) knowledge base module.



Intervention Procedure

Participants in the intervention group learned about osteoporosis using the OPBot, a chatbot-based educational tool. The intervention was initiated during hospitalization, and participants were able to continue using the OPBot after discharge. Specifically, after osteoporosis was confirmed based on examination results (usually within 1-2 days after hospital admission), nurses informed participants of the diagnosis at their bedside and provided instructions on how to log in to and use the OPBot on their own smartphones.

During hospitalization, participants were encouraged to use OPBot as their primary source of SME. Nurses were available for consultation during working hours; however, their role was limited to addressing issues not resolved by OPBot. In addition, a specialist nurse and a case manager monitored the OPBot

backend once daily to identify unresolved queries or potential risks flagged by the system. In-person intervention was provided only when necessary, based on predefined criteria (eg, unanswered questions, incorrect responses, or signs of misunderstanding), and was delivered either face-to-face or online. This targeted support was intended to supplement, rather than replace, the chatbot's function. Before discharge, participants underwent a second knowledge assessment.

Comparison

The control group received traditional health education, which included face-to-face communication and written materials. After participants were enrolled, nurses provided health education and distributed an osteoporosis health handbook for participants to read. During hospitalization, participants in the control group could consult nurses during working hours

regarding osteoporosis treatment and home care. All consultations were initiated by participants and conducted routinely, without systematic monitoring or proactive follow-up by nurses. No digital tools or automated support systems were used in this group. Before discharge, participants underwent a second knowledge assessment.

Data Collection

Data were collected to assess the feasibility and to explore the potential effectiveness of the intervention.

Feasibility Outcomes

Feasibility was assessed from 2 aspects: nurse-reported working time and the reliability of OPBot responses. Nurse-reported working time was assessed by asking nurses to self-record their working hours during each intervention session. In the intervention group, nurses' time included health education, consultation, and operational guidance. In the control group, nurses' time included health education and consultation. Total working time was calculated by summing the recorded durations across sessions.

The reliability of OPBot responses was evaluated during the study. Participants asked questions and received corresponding answers through the question-and-answer module of OPBot. question-and-answer records were extracted from the OPBot backend for evaluation. Two assessors (a specialist nurse and the case manager) independently reviewed the responses. The appropriateness of each response was rated using a predefined 5-point Likert scale: "highly reliable," "mostly reliable," "partly reliable," "unreliable," and "unable to answer." Interrater reliability between the 2 assessors was assessed using the Cohen kappa coefficient, and agreement was interpreted according to the criteria of Landis and Koch [37]. When discrepancies occurred between the 2 assessors, the responses were further reviewed by the director of the orthopedic department, and a consensus rating was reached.

Effectiveness Outcomes

Effectiveness was explored in terms of participants' osteoporosis-related knowledge and self-management adherence.

Knowledge scores were assessed using the Osteoporosis Knowledge Assessment Tool (OKAT). The Chinese version of the OKAT has demonstrated good reliability and validity in previous studies [38]. The OKAT is provided in [Multimedia Appendix 2](#). The OKAT consists of 20 items and the overall Cronbach α was 0.821. All questionnaires were administered and collected by researcher C and researcher D. Before implementation of the intervention or control group measures, the first osteoporosis knowledge score assessment was conducted as the baseline value. On the day of discharge, participants underwent a second knowledge assessment.

Self-management adherence was assessed using a self-developed questionnaire during telephone follow-ups at 1, 3, and 6 months after discharge. The questionnaire had a total score of 100 and consisted of 4 domains: calcium supplementation, consumption of calcium-rich foods, sunlight exposure, and exercise frequency. Each domain had a maximum score of 25 and was rated on a 3-point scale (0=none, 12.5=sometimes, and

25=always). The total adherence score was calculated as the sum of the scores across the 4 domains, with higher scores indicating better adherence. The items were developed based on osteoporosis self-management recommendations used in clinical practice. The full questionnaire is provided in [Multimedia Appendix 3](#).

Statistical Analysis

All statistical analyses were performed using R software (version 4.4.3; R Foundation for Statistical Computing). Continuous variables were summarized as mean (SD) or median and IQR, as appropriate, and compared between groups using independent *t* tests or Mann-Whitney *U* tests. Categorical variables were analyzed using chi-square tests or Fisher exact tests.

For adherence outcomes measured at 1, 3, and 6 months, mixed-effects models were used to account for within-subject correlations arising from repeated measurements over time. Depending on the nature of the outcome, logistic mixed-effects models were applied for binary outcomes, ordinal mixed-effects models for ordered categorical outcomes, and linear mixed-effects models for continuous outcomes. In all mixed-effects models, fixed effects included group (intervention vs control), time (1, 3, and 6 months), and their interaction (group \times time), along with baseline covariates (sex, age, and education level). A random intercept for each participant was included to account for repeated measurements within individuals. The group \times time interaction term was used to evaluate whether the effect of the intervention varied over time. To address multiplicity across adherence outcomes, calcium supplement intake was prespecified as the primary adherence outcome and was analyzed without adjustment for multiple testing. The remaining adherence outcomes, including calcium-rich food consumption, sun exposure, weekly exercise frequency, and total adherence score, were treated as secondary outcomes. Holm correction was applied to adjust for multiple comparisons among the secondary outcomes, and adjusted *P* values are reported accordingly. A 2-sided *P*<.05 was considered statistically significant.

Results

Study Population and Characteristics

A total of 643 participants were screened before the study, and 15.6% (100/643) participants who met the study criteria and volunteered to participate were included and randomized into the intervention (n=50, 50%) or control (n=50, 50%) group. Ultimately, 45% (45/100) patients in the intervention group and 43 (43/100) patients in the control group completed the trial, resulting in a dropout rate of 12% (12/100). This corresponds to an overall missing data rate of 12%, as missing data arose from participant dropout. In the experimental group, 2% (2/100) participants did not complete the second knowledge assessment, and 3% (3/100) participants were lost to follow-up. In the control group, 1% (1/100) participant did not complete the second knowledge assessment, 2% (2/100) participants died, and 4% (4/100) participants were lost to follow-up ([Figure 2](#)). At baseline, no significant differences were observed in sex distribution, education level, and age between the 2 groups ([Table 1](#)). In addition, comparisons between included and

excluded participants showed no statistically significant differences in baseline characteristics (Table S1 in [Multimedia Appendix 4](#)), suggesting that attrition was not associated with the measured baseline variables.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart of participant enrollment, allocation, and adherence. OKAT: Osteoporosis Knowledge Assessment Tool.

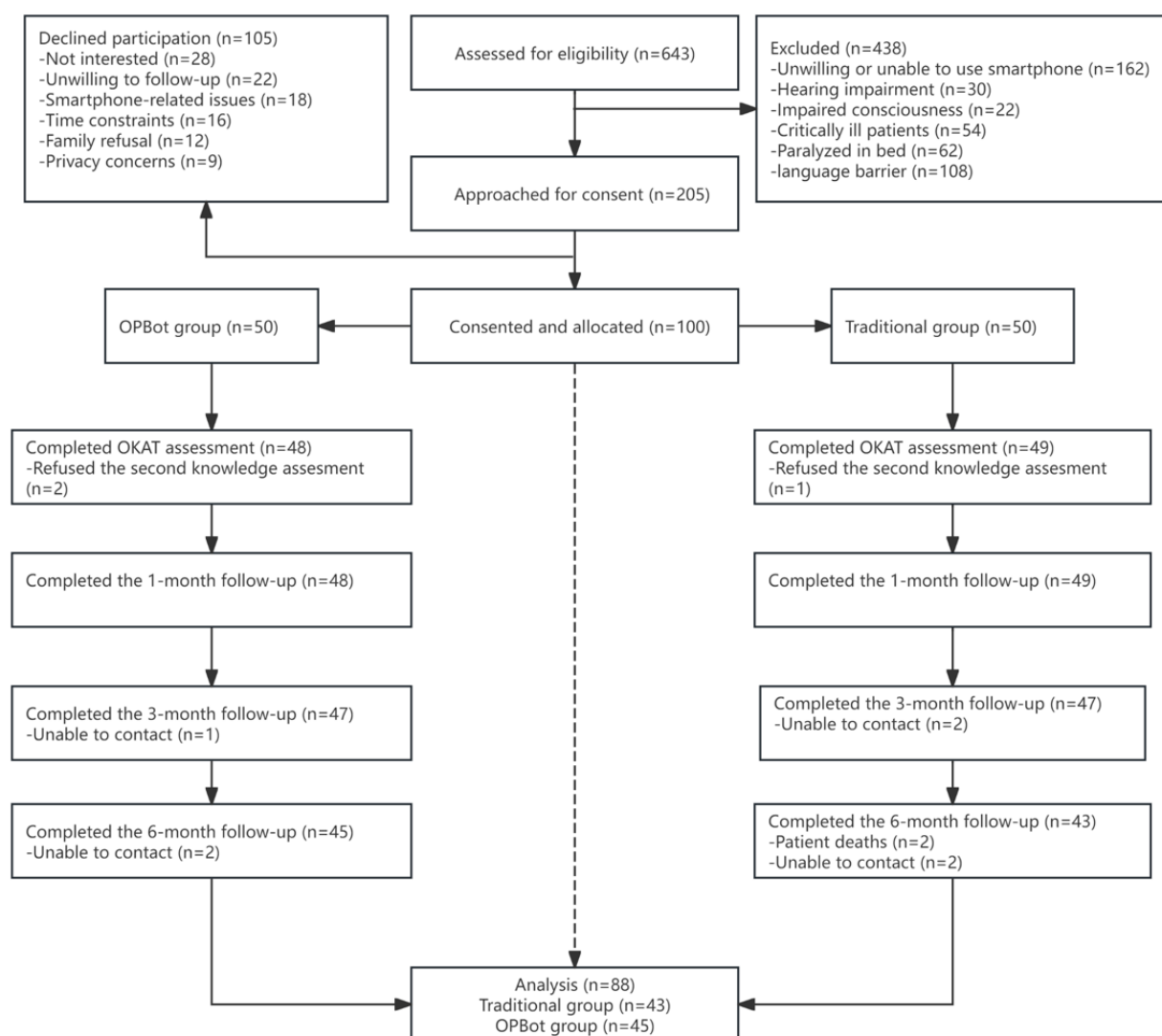


Table 1. Baseline characteristics of participants by study group (intervention vs control). Data are presented for participants with complete follow-up data.

	Intervention (n=45)	Control (n=43)	P value
Sex, n (%)			.13
Female	29 (64.4)	34 (79.1)	
Male	16 (35.6)	9 (20.9)	
Age (years), mean (SD)	68.58 (9.4)	72.56 (9.8)	.06
Education, n (%)			.84
Illiteracy	9 (20)	7 (16.3)	
Primary or secondary school	35 (77.8)	34 (79.1)	
College or undergraduate	1 (2.2)	2 (4.6)	

Feasibility Outcomes

Nurse-Reported Working Time

Nurse-reported working time was significantly lower in the

intervention group compared with the control group (median 5.00, IQR 2.00-17.00 vs median 23.00, IQR 20.00-25.00; $P<.001$; [Table 2](#)).

Table 2. Knowledge scores, time interval between assessments, and nurse-reported working time by study group. Relevant findings are described in the corresponding sections.

	Intervention (n=45), median (IQR)	Control (n=43), median (IQR)	<i>P</i> value
Knowledge score			
A1 ^a	55.00 (42.00-70.00)	53.00 (39.50-66.50)	.66
A2 ^b	80.00 (70.00-89.00)	75.00 (65.50-80.00)	.01
Time interval between A1 and A2 (days)	5.0 (3.0-7.0)	5.0 (4.0-6.0)	.43
Working hours			
Nurse-reported working time (min)	5.00 (2.00-17.00)	23.00 (20.00-25.00)	<.001

^aA1: the baseline of knowledge score about osteoporosis.

^bA2: postintervention osteoporosis knowledge score.

Reliability of OPBot

In the question-and-answer module, a total of 85 questions and corresponding responses were recorded. Initial disagreement was identified in 3 records and was resolved by consensus. Interrater agreement was almost perfect (Cohen $\kappa=0.83$). Final ratings showed that 76 (89.4%) records were highly reliable, 1 (1.2%) record was mostly reliable, 1 (1.2%) record was partially reliable, 1 (1.2%) record was unreliable, and 6 (7.1%) records were classified as unable to respond.

Effectiveness Outcomes

Knowledge Score

Regarding knowledge scores, no significant difference was observed between the 2 groups at baseline (A1; median 55.0, IQR 42.0-70.0 vs median 53.0, IQR 39.5-66.5; $P=.66$). After the intervention, the intervention group demonstrated significantly higher knowledge scores (A2) compared with the control group (median 80.0, IQR 70.0-89.0 vs median 75.0, IQR 65.5-80.0; $P=.01$). The median time interval between the 2 knowledge assessments was 5.0 days in both groups (intervention: IQR 3.0-7.0; control: IQR 4.0-6.0; $P=.43$; [Table 2](#)).

Self-Management Adherence Outcomes

Adherence outcomes are presented in [Table 3](#). For the primary outcome, calcium supplement intake, the main effect of group was not statistically significant (odds ratio [OR] 0.37, 95% CI 0.10-1.37; $P=.14$). However, a significant group \times time interaction was observed (OR 1.49, 95% CI 1.08-2.06; $P=.02$), indicating that changes in supplement intake over time differed between the intervention and control groups. A significant main effect of time was also identified for calcium supplement intake (OR 0.75, 95% CI 0.61-0.93; $P=.01$).

For the secondary outcomes, the intervention group showed higher odds of consuming calcium-rich foods than the control group (OR 2.87, 95% CI 1.04-7.89; nominal $P=.04$). However, this association did not remain statistically significant after Holm correction for multiple comparisons. No significant group differences were observed for sun exposure, weekly exercise frequency, or total adherence scores. A significant time effect was observed for sun exposure (OR 0.97, 95% CI 0.94-0.99; $P=.03$), suggesting that sun exposure changed over time across both groups. No significant group \times time interactions were identified for any secondary outcomes after Holm correction.

Table 3. Results of mixed-effects models for adherence outcomes over time.

	Calcium supplement intake		Calcium-rich foods consumption		Sun exposure		Weekly frequency of exercise		Total scores	
	OR ^a (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	β (95% CI)	P value
Group										
Control	Reference		Reference		Reference		Reference		Reference	
Intervention	0.37 (0.10-1.37)	.14	2.87 (1.04-7.89)	.04	0.99 (0.39-2.54)	.99	1.20 (0.42-3.42)	.73	1.88 (-8.16-11.91)	.71
Sex										
Female	Reference		Reference		Reference		Reference		Reference	
Male	1.20 (0.52-2.77)	.66	1.46 (0.63-3.37)	.38	1.15 (0.59-2.25)	.68	0.48 (0.23-0.99)	.05	-1.06 (-8.39-6.27)	.78
Age	0.99 (0.95-1.03)	.67	1.03 (0.99-1.07)	.18	0.97 (0.94-0.99)	.03	0.98 (0.95-1.02)	.36	-0.08 (-0.43-0.27)	.65
Education	0.94 (0.41-2.20)	.89	1.83 (0.83-4.07)	.14	0.79 (0.40-1.54)	.48	0.97 (0.44-2.13)	.94	1.54 (-5.95-9.03)	.69
Time	0.75 (0.61-0.93)	.01	1.02 (0.85-1.21)	.85	0.89 (0.76-1.05)	.18	1.05 (0.87-1.26)	.64	-1.49 (-3.29-0.31)	.11
Group×time ^b	1.49 (1.08-2.06)	.02	0.89 (0.70-1.15)	.79	1.17 (0.93-1.48)	.56	1.12 (0.85-1.48)	.79	2.25 (-0.27-4.76)	.33

^aOR: odds ratio.

^bGroup×time: interaction between group and time.

Discussion

Principal Findings

This formative randomized controlled trial aimed to preliminarily explore the feasibility and potential effects of a chatbot-based SME tool (OPBot) among patients with osteoporosis. For feasibility, nurses in the intervention group spent significantly less time on SME than those in the control group, indicating that OPBot may have the potential to reduce the workload associated with patient education. In addition, the reliability evaluation of the question-and-answer module showed a high level of interrater agreement, suggesting that OPBot responses demonstrated a relatively high level of reliability in a clinical context.

For effectiveness outcomes, the intervention group had significantly higher postintervention knowledge scores compared with the control group. In this study, the time interval between baseline and postintervention assessments was relatively short (median 5 days) and did not differ significantly between groups, which may reduce the likelihood of differential bias. Nevertheless, because the OPBot intervention included a module developed based on the same knowledge assessment tool, participants in the intervention group may have been more exposed to similar content and therefore more likely to recall or recognize test items. This may have led to an overestimation of the intervention effect on knowledge outcomes.

In addition, the intervention group demonstrated a more favorable trajectory over time for calcium supplement intake, as reflected by a significant group×time interaction. This suggests that the pattern of supplement use evolved differently

between groups across the follow-up period, rather than showing a consistent difference at individual time points. The intervention group also showed higher odds of consuming calcium-rich foods across time points, although this association did not remain significant after adjustment for multiple comparisons. However, no significant effects were observed for sun exposure, exercise, or total adherence scores, suggesting that the impact of the intervention on adherence was behavior specific rather than uniform across all domains. Overall, these findings indicate that chatbot-based health education may influence certain aspects of self-management adherence, although its effects may vary depending on the specific behavior.

Comparison With Prior Work

Previous research has shown that LLM-driven chatbots can provide high-quality and easily understandable health information, demonstrating good potential in improving patient understanding and information access [23,25]. Our findings are consistent with these studies, showing that LLM chatbots may have a positive role in improving self-management knowledge.

Different adherence patterns emerged across various behaviors. Specifically, the changes in calcium supplement intake between the intervention and control groups (group×time interaction: OR 1.49, 95% CI 1.08-2.06; nominal $P=.02$), although this difference was not statistically significant after Holm correction. The intervention group consistently demonstrated higher adherence to calcium-rich foods at all time points. However, no significant differences were observed in other adherence behaviors or overall adherence scores, suggesting that the effectiveness of educational interventions is behavior specific [39]. Comparatively, behaviors such as supplement intake and

dietary choices were more susceptible to the influence of chatbot-assisted education, possibly due to their easier integration into daily life, lower implementation costs, and greater consistency with existing health perceptions. For example, in some cultural contexts (including southern China), diet plays a crucial role in health management, which may further promote the adoption of relevant health recommendations. From a practical perspective, these results suggest that chatbot-based health education interventions should consider both the behavioral characteristics and cultural context of the target population when designing them. For behaviors more easily driven by information (such as diet and supplement intake), enhanced personalized recommendations and continuous feedback can improve effectiveness, while for behaviors more influenced by environment or habits (such as exercise or sun exposure), more complex behavioral support strategies may be necessary. Overall, continued access to educational support may help improve specific compliance behaviors, but its impact on overall self-management remains behavior dependent.

Moreover, although general-purpose LLM-based chatbots (eg, ChatGPT; OpenAI) have demonstrated high levels of readability and interactivity in providing general health information [23,25,28], they may still produce responses that are incomplete, insufficiently specific, or potentially inaccurate when addressing disease-specific questions, particularly in clinical contexts [23-26,40]. In contrast, the OPBot in this study was optimized for a specific disease domain, integrating a real clinical knowledge base and undergoing targeted training, enabling it to improve the professionalism and relevance of its answers while maintaining a good interactive experience. This “domain-customized” design may, to some extent, compensate for the shortcomings of general-purpose models regarding professional depth, thus making it more suitable for specific scenarios of chronic disease SME.

Implications

Furthermore, integrating OPBot into the health care system, especially in-hospital clinical education, may help support SME delivery. Although there have been studies on LLM-based chatbots in education, there is still a lack of research on their use within hospital settings [17,25,27]. Typically, SME starts when a patient is admitted to the hospital and runs through the entire treatment and follow-up process. However, the introduction of OPBot may help reshape this traditional process. On the basis of our study, OPBot could serve as a knowledge coach, mainly responsible for providing patients with general osteoporosis knowledge and answering specialized questions. Although LLM-powered chatbots have shown great potential in personalized guidance [25,41], the accuracy and reliability of information are still key issues that must be addressed, especially in the medical field, where any misinformation may have a negative impact on patients. Therefore, current chatbots cannot completely replace the role of human providers [42]. In

actual applications, neither nurses nor patients rely on OPBot alone but rather regard it as an auxiliary tool.

Furthermore, Beg [43] emphasized that artificial intelligence (AI) research should establish clear standards for ethical review and transparency to safeguard participants’ rights and mitigate potential risks. While AI shows great promise in emotion recognition, dynamic feedback, and remote intervention, it also faces challenges related to trust, ethical responsibility, and role coordination. Accordingly, this study strengthened data privacy protection, information traceability, and mechanisms for human oversight. For instance, nurses reviewed OPBot responses daily and proactively contacted patients when necessary to provide personalized guidance, reflecting the ethical principle of human-AI collaboration.

Moreover, although chatbots can provide real-time feedback and personalized learning based on patients’ needs, a lack of transparency in algorithmic logic and information sources may lead to bias or misinformation. As Beg and Verma [44] further highlighted, personalization and clinical interpretability are essential in AI-driven digital interventions. Therefore, this study implemented safety mechanisms within OPBot. For example, any inquiry related to medication or treatment automatically triggered the response “Please consult your doctor.” This design ensured that patients made medical decisions under professional supervision, embodying the importance of ethical risk control.

Limitations

However, this study still has certain limitations. First, the reliance on self-reported adherence data is prone to recall and social desirability bias. In addition, although no significant baseline differences were observed between included and excluded participants (Table S1 in [Multimedia Appendix 4](#)), the potential impact of missing data and attrition bias cannot be completely ruled out. Second, the participants were recruited from a single tertiary hospital in China, which may limit the generalizability of the findings. Future studies should conduct multicenter, large-sample randomized controlled trials to provide higher-quality evidence. Extending the follow-up duration in future research would allow for a more comprehensive evaluation of the long-term effectiveness of chatbot-based SME, such as its impact on patients’ bone density.

Conclusions

OPBot may have the potential to improve patients’ self-management knowledge of their disease. In health care settings with limited human resources, leveraging educational chatbots to share part of the educational workload and provide patients with basic self-management guidance may be a feasible strategy. Furthermore, to provide better personalized guidance to patients, continuously expanding the chatbot’s corpus is essential to improving its accuracy and credibility. Finally, this study offers valuable insights for SME for other chronic diseases, although further large-scale and long-term studies are needed to confirm these findings.

Acknowledgments

During manuscript preparation, the first author used a generative artificial intelligence (AI) tool (ChatGPT; OpenAI) to assist with proofreading and language editing of selected sections of the text. According to the Generative AI Delegation Taxonomy

(GAIDeT) framework (2025), such tools may be used for editorial support under full human supervision. All core aspects of this study, including study design, conceptual framework development, data analysis, and interpretation of results, were conducted independently by the authors. No part of the manuscript was generated autonomously without human oversight, and all critical academic decisions and final revisions were made by the authors. The AI tool was not listed as an author and bears no responsibility for the content, analysis, or conclusions of this study.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to ethical and privacy considerations but are available from the corresponding author on reasonable request.

Funding

This work was supported by the 2025 Clinical and Basic Science and Technology Innovation Special Program (Key Project) of Guangdong Medical University (grant GDMULCJC202500).

Authors' Contributions

Conceptualization: JH, XX, XD

Data curation: JH

Formal analysis: HX

Investigation: XX, CH, WP, SP

Methodology: JH, HX

Software: WS

Supervision: XD

Validation: JH, XX

Writing—original draft: JH

Writing—review and editing: HX, XD

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was not registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered formative, guiding the development of the tool. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1

The context and design process of OPBot.

[\[DOCX File , 23 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Osteoporosis knowledge assessment tool (Chinese version).

[\[DOCX File , 24 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Osteoporosis follow-up assessment form.

[\[DOCX File , 22 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Comparison of baseline characteristics between included and excluded participants.

[\[DOCX File , 13 KB-Multimedia Appendix 4\]](#)

References

1. Search results. World Health Organization. URL: <https://www.who.int/home/search-results?indexCatalogue=genericsearchindex1&searchQuery=death&wordsMode=AnyWord#gsc.tab=0> [accessed 2025-07-31]

2. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. *Annu Rev Public Health*. Apr 01, 2019;40(1):127-146. [FREE Full text] [doi: [10.1146/annurev-publhealth-040218-044008](https://doi.org/10.1146/annurev-publhealth-040218-044008)] [Medline: [30601717](https://pubmed.ncbi.nlm.nih.gov/30601717/)]
3. Hassan NM, Win KT, Hyland P. Exploring design features and benefits of Online Patient Education (OPE) sites for chronic diseases. In: *Proceedings of the 46th Hawaii International Conference on System Sciences*. 2013. Presented at: HICSS 2013; Jan 7-10, 2013; Wailea, HI. [doi: [10.1109/HICSS.2013.215](https://doi.org/10.1109/HICSS.2013.215)]
4. Weiss BD. *Health Literacy: A Manual for Clinicians*. Chicago, IL. American Medical Association; 2003.
5. Chodosh J, Morton SC, Mojica W, Maglione M, Suttrop MJ, Hilton L, et al. Meta-analysis: chronic disease self-management programs for older adults. *Ann Intern Med*. Sep 20, 2005;143(6):427-438. [doi: [10.7326/0003-4819-143-6-200509200-00007](https://doi.org/10.7326/0003-4819-143-6-200509200-00007)] [Medline: [16172441](https://pubmed.ncbi.nlm.nih.gov/16172441/)]
6. Farley H. Promoting self-efficacy in patients with chronic disease beyond traditional education: a literature review. *Nurs Open*. Oct 20, 2019;7(1):30-41. [FREE Full text] [doi: [10.1002/nop2.382](https://doi.org/10.1002/nop2.382)] [Medline: [31871689](https://pubmed.ncbi.nlm.nih.gov/31871689/)]
7. Health workforce. World Health Organization. URL: <https://www.who.int/health-topics/health-workforce> [accessed 2026-03-24]
8. Stelfson M, Paige SR, Chaney BH, Chaney JD. Evolving role of social media in health promotion: updated responsibilities for health education specialists. *Int J Environ Res Public Health*. Feb 12, 2020;17(4):1153. [FREE Full text] [doi: [10.3390/ijerph17041153](https://doi.org/10.3390/ijerph17041153)] [Medline: [32059561](https://pubmed.ncbi.nlm.nih.gov/32059561/)]
9. Borghouts J, Eikev E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *J Med Internet Res*. Mar 24, 2021;23(3):e24387. [FREE Full text] [doi: [10.2196/24387](https://doi.org/10.2196/24387)] [Medline: [33759801](https://pubmed.ncbi.nlm.nih.gov/33759801/)]
10. Atiyah A, Jusoh S, Alghanim F. Evaluation of the naturalness of chatbot applications. In: *Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology*. 2019. Presented at: JEEIT 2019; Apr 9-11, 2019; Amman, Jordan. [doi: [10.1109/jeeit.2019.8717455](https://doi.org/10.1109/jeeit.2019.8717455)]
11. Svenningsson N, Faraon M. Artificial intelligence in conversational agents: a study of factors related to perceived humanness in chatbots. In: *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*. 2019. Presented at: AICCC '19; Dec 21-23, 2019; Kobe, Japan. [doi: [10.1145/3375959.3375973](https://doi.org/10.1145/3375959.3375973)]
12. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. Sep 01, 2018;25(9):1248-1258. [FREE Full text] [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
13. Adamopoulou E, Moussiades L. An overview of chatbot technology. In: *Proceedings of the 16th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations*. 2020. Presented at: AIAI 2020; Jun 5-7, 2020; Neos Marmaras, Greece. [doi: [10.1007/978-3-030-49186-4_31](https://doi.org/10.1007/978-3-030-49186-4_31)]
14. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digit Med*. Dec 19, 2023;6(1):236. [FREE Full text] [doi: [10.1038/s41746-023-00979-5](https://doi.org/10.1038/s41746-023-00979-5)] [Medline: [38114588](https://pubmed.ncbi.nlm.nih.gov/38114588/)]
15. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. *Information*. Jan 15, 2022;13(1):41. [doi: [10.3390/info13010041](https://doi.org/10.3390/info13010041)]
16. Kim HW, Shin DH, Kim J, Lee GH, Cho JW. Assessing the performance of ChatGPT's responses to questions related to epilepsy: a cross-sectional study on natural language processing and medical information retrieval. *Seizure*. Jan 2024;114:1-8. [FREE Full text] [doi: [10.1016/j.seizure.2023.11.013](https://doi.org/10.1016/j.seizure.2023.11.013)] [Medline: [38007922](https://pubmed.ncbi.nlm.nih.gov/38007922/)]
17. Aydin S, Karabacak M, Vlachos V, Margetis K. Navigating the potential and pitfalls of large language models in patient-centered medication guidance and self-decision support. *Front Med (Lausanne)*. Jan 23, 2025;12:1527864. [FREE Full text] [doi: [10.3389/fmed.2025.1527864](https://doi.org/10.3389/fmed.2025.1527864)] [Medline: [39917061](https://pubmed.ncbi.nlm.nih.gov/39917061/)]
18. Zheng X, Li Z, Gui X, Luo Y. Customizing emotional support: how do individuals construct and interact with LLM-powered chatbots. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 2025. Presented at: CHI '25; Apr 26-May 1, 2025; Yokohama, Japan. [doi: [10.1145/3706598.3713453](https://doi.org/10.1145/3706598.3713453)]
19. Wang Y, Wang Y, Xiao Y, Escamilla L, Augustine B, Crace K, et al. Evaluating an LLM-powered chatbot for cognitive restructuring: insights from mental health professionals. *arXiv*. Preprint posted online on January 26, 2025. [doi: [10.48550/arXiv.2501.15599](https://doi.org/10.48550/arXiv.2501.15599)]
20. Hao Y, Holmes J, Waddle MR, Davis BJ, Yu NY, Vickers KS, et al. Personalizing prostate cancer education for patients using an EHR-integrated LLM agent. *NPJ Digit Med*. Dec 18, 2025;8(1):770. [FREE Full text] [doi: [10.1038/s41746-025-02166-0](https://doi.org/10.1038/s41746-025-02166-0)] [Medline: [41413170](https://pubmed.ncbi.nlm.nih.gov/41413170/)]
21. Ramjee P, Sachdeva B, Golechha S, Kulkarni S, Fulari G, Murali K, et al. CataractBot: an LLM-powered expert-in-the-loop chatbot for cataract patients. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. Jun 18, 2025;9(2):1-31. [doi: [10.1145/3729479](https://doi.org/10.1145/3729479)]
22. Lee A, Moon S, Jhon M, Kim JW, Kim DK, Kim JE. Comparative study on the performance of LLM-based psychological counseling chatbots via prompt engineering techniques. In: *Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine*. 2024. Presented at: BIBM 2024; Dec 3-6, 2024; Lisbon, Portugal. [doi: [10.1109/bibm62325.2024.10822158](https://doi.org/10.1109/bibm62325.2024.10822158)]

23. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* Jun 01, 2023;183(6):589-596. [FREE Full text] [doi: [10.1001/jamainternmed.2023.1838](https://doi.org/10.1001/jamainternmed.2023.1838)] [Medline: [37115527](https://pubmed.ncbi.nlm.nih.gov/37115527/)]
24. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health.* Feb 9, 2023;2(2):e0000198. [FREE Full text] [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
25. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: a systematic review and meta-analysis. *J Biomed Inform.* Mar 2024;151:104620. [FREE Full text] [doi: [10.1016/j.jbi.2024.104620](https://doi.org/10.1016/j.jbi.2024.104620)] [Medline: [38462064](https://pubmed.ncbi.nlm.nih.gov/38462064/)]
26. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst.* Mar 04, 2023;47(1):33. [FREE Full text] [doi: [10.1007/s10916-023-01925-4](https://doi.org/10.1007/s10916-023-01925-4)] [Medline: [36869927](https://pubmed.ncbi.nlm.nih.gov/36869927/)]
27. Huo B, Boyle A, Marfo N, Tangamornsuksan W, Steen JP, McKechnie T, et al. Large language models for chatbot health advice studies: a systematic review. *JAMA Netw Open.* Feb 03, 2025;8(2):e2457879. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.57879](https://doi.org/10.1001/jamanetworkopen.2024.57879)] [Medline: [39903463](https://pubmed.ncbi.nlm.nih.gov/39903463/)]
28. Bedi S, Liu Y, Orr-Ewing L, Dash D, Koyejo S, Callahan A, et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA.* Jan 28, 2025;333(4):319-328. [doi: [10.1001/jama.2024.21700](https://doi.org/10.1001/jama.2024.21700)] [Medline: [39405325](https://pubmed.ncbi.nlm.nih.gov/39405325/)]
29. Agrawal M, Chen IY, Gulamali F, Joshi S. The evaluation illusion of large language models in medicine. *NPJ Digit Med.* Oct 07, 2025;8(1):600. [FREE Full text] [doi: [10.1038/s41746-025-01963-x](https://doi.org/10.1038/s41746-025-01963-x)] [Medline: [41057566](https://pubmed.ncbi.nlm.nih.gov/41057566/)]
30. Sabri SA, Chavarria JC, Ackert-Bicknell C, Swanson C, Burger E. Osteoporosis: an update on screening, diagnosis, evaluation, and treatment. *Orthopedics.* 2023;46(1):e20-e26. [FREE Full text] [doi: [10.3928/01477447-20220719-03](https://doi.org/10.3928/01477447-20220719-03)] [Medline: [35876780](https://pubmed.ncbi.nlm.nih.gov/35876780/)]
31. Cappelle SI, Ramon I, Dekelver C, Rozenberg S, Baleanu F, Karmali R, et al. Distribution of clinical risk factors for fracture in a Brussels cohort of postmenopausal women: the FRISBEE study and comparison with other major cohort studies. *Maturitas.* Dec 2017;106:1-7. [doi: [10.1016/j.maturitas.2017.08.010](https://doi.org/10.1016/j.maturitas.2017.08.010)] [Medline: [29150162](https://pubmed.ncbi.nlm.nih.gov/29150162/)]
32. Borgström F, Karlsson L, Ortsäter G, Norton N, Halbout P, Cooper C, et al. Fragility fractures in Europe: burden, management and opportunities. *Arch Osteoporos.* Apr 19, 2020;15(1):59. [FREE Full text] [doi: [10.1007/s11657-020-0706-y](https://doi.org/10.1007/s11657-020-0706-y)] [Medline: [32306163](https://pubmed.ncbi.nlm.nih.gov/32306163/)]
33. Eisman JA, Bogoch ER, Dell R, Harrington JT, McKinney REJ, McLellan A, et al. Making the first fracture the last fracture: ASBMR task force report on secondary fracture prevention. *J Bone Miner Res.* Oct 2012;27(10):2039-2046. [FREE Full text] [doi: [10.1002/jbmr.1698](https://doi.org/10.1002/jbmr.1698)] [Medline: [22836222](https://pubmed.ncbi.nlm.nih.gov/22836222/)]
34. Barker KL, Toye F, Lowe CJ. A qualitative systematic review of patients' experience of osteoporosis using meta-ethnography. *Arch Osteoporos.* Dec 2016;11(1):33. [FREE Full text] [doi: [10.1007/s11657-016-0286-z](https://doi.org/10.1007/s11657-016-0286-z)] [Medline: [27739032](https://pubmed.ncbi.nlm.nih.gov/27739032/)]
35. Huang J, Ding X. Self-management support (SMS) in transition: the case of osteoporosis management support in a Chinese hospital. *Comput Support Coop Work.* Nov 18, 2024;34:1-39. [doi: [10.1007/s10606-024-09503-z](https://doi.org/10.1007/s10606-024-09503-z)]
36. Gendler PE, Coviak CP, Martin JT, Kim KK, Dankers JK, Barclay JM, et al. Revision of the osteoporosis knowledge test: reliability and validity. *West J Nurs Res.* Dec 2015;37(12):1623-1643. [doi: [10.1177/0193945914537565](https://doi.org/10.1177/0193945914537565)] [Medline: [24923463](https://pubmed.ncbi.nlm.nih.gov/24923463/)]
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* Mar 1977;33(1):159-174. [doi: [10.2307/2529310](https://doi.org/10.2307/2529310)]
38. Guo Q, Guo M, Zhang L, Zhao P, Guo L, Gao H, et al. Reliability and validity of Chinese version of the Osteoporosis Knowledge Assessment Tool in the elderly. *Chin J Pract Nurs.* 2015;(36):512-515. [FREE Full text]
39. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med.* Nov 2016;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
40. Shiferaw MW, Zheng T, Winter A, Mike LA, Chan LN. Assessing the accuracy and quality of artificial intelligence (AI) chatbot-generated responses in making patient-specific drug-therapy and healthcare-related decisions. *BMC Med Inform Decis Mak.* Dec 24, 2024;24(1):404. [FREE Full text] [doi: [10.1186/s12911-024-02824-5](https://doi.org/10.1186/s12911-024-02824-5)] [Medline: [39719573](https://pubmed.ncbi.nlm.nih.gov/39719573/)]
41. Abbasian M, Azimi I, Rahmani AM, Jain R. Conversational health agents: a personalized large language model-powered agent framework. *JAMIA Open.* Jul 06, 2025;8(4):ooaf067. [doi: [10.1093/jamiaopen/ooaf067](https://doi.org/10.1093/jamiaopen/ooaf067)] [Medline: [40620478](https://pubmed.ncbi.nlm.nih.gov/40620478/)]
42. Bean AM, Payne RE, Parsons G, Kirk HR, Ciro J, Mosquera-Gómez R, et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. *Nat Med.* Feb 2026;32(2):609-615. [doi: [10.1038/s41591-025-04074-y](https://doi.org/10.1038/s41591-025-04074-y)] [Medline: [41663592](https://pubmed.ncbi.nlm.nih.gov/41663592/)]
43. Beg MJ. Qualitative methods in mental health research: standards for ethical inquiry, research practice, and peer review. *Indian J Psychol Med.* Aug 10, 2025;02537176251363817. [doi: [10.1177/02537176251363817](https://doi.org/10.1177/02537176251363817)] [Medline: [40799658](https://pubmed.ncbi.nlm.nih.gov/40799658/)]
44. Beg MJ, Verma MK. Exploring the potential and challenges of digital and AI-driven psychotherapy for ADHD, OCD, schizophrenia, and substance use disorders: a comprehensive narrative review. *Indian J Psychol Med.* Dec 14, 2024;02537176241300569. [doi: [10.1177/02537176241300569](https://doi.org/10.1177/02537176241300569)] [Medline: [39677516](https://pubmed.ncbi.nlm.nih.gov/39677516/)]

Abbreviations

AI: artificial intelligence
IOF: International Osteoporosis Foundation
LLM: large language model
OKAT: Osteoporosis Knowledge Assessment Tool
OR: odds ratio
SME: self-management education

Edited by S Law; submitted 08.Oct.2025; peer-reviewed by J Odeniya, AJ Umeano; comments to author 13.Mar.2026; accepted 12.May.2026; published 02.Jun.2026

Please cite as:

Huang J, Xin X, He C, Xiong H, Pang W, Pang S, Sun W, Ding X

Preliminary Evaluation of a Large Language Model–Powered Chatbot for Osteoporosis Self-Management Education: Formative Randomized Controlled Trial

JMIR Form Res 2026;10:e85475

URL: <https://formative.jmir.org/2026/1/e85475>

doi: [10.2196/85475](https://doi.org/10.2196/85475)

PMID:

©Jinling Huang, Xiaolian Xin, Chunyan He, Haoyan Xiong, Wenjun Pang, Shaobo Pang, Wei Sun, Xianghua Ding. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 02.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.