Psychogeriatrics

ORIGINAL ARTICLE OPEN ACCESS

A Conversational Robot for Cognitively Impaired Older People Who Live Alone: An Exploratory Feasibility Study

Yuma Nagata[1,2] | Yuto Satake[1,3] | Ryuji Yamazaki[4,5] | Shuichi Nishio[4] | Maki Suzuki[6] | Hideki Kanemoto[1,7] | Miyae Yamakawa[8] | David Figueroa[9,10] | Elie Maalouly[10,11] | Hiroshi Ishiguro[4,10] | Manabu Ikeda[1]

[1]Department of Psychiatry, The University of Osaka Graduate School of Medicine, Suita, Osaka, Japan | [2]Division of Occupational Therapy, Department of Rehabilitation, Faculty of Health Science, Naragakuen University, Nara, Japan | [3]Division of Psychiatry, University College London, London, UK | [4]Symbiotic Intelligent Systems Research Center, Institute for Open and Transdisciplinary Research Initiatives, The University of Osaka, Suita, Osaka, Japan | [5]Education Center for Regional Revitalization/Faculty of Global Policy Management and Communication, Yamanashi Prefectural University, Yamanashi, Japan | [6]Department of Behavioral Neurology and Neuropsychiatry, The University of Osaka United Graduate School of Child Development, Suita, Osaka, Japan | [7]Health and Counseling Center, The University of Osaka, Toyonaka, Osaka, Japan | [8]Division of Health Sciences, Graduate School of Medicine, The University of Osaka, Suita, Osaka, Japan | [9]Department of Materials and Production, Aalborg University, Aalborg, Denmark | [10]Intelligent Robotics Laboratory, Department of Systems Innovation, Graduate School of Engineering Science, The University of Osaka, Suita, Osaka, Japan | [11]Department of Neuroscience and Reproductive and Odontostomatological Sciences, University of Naples Federico II, Naples, Italy

Correspondence: Yuto Satake (y.satake@psy.med.osaka-u.ac.jp)

Received: 12 June 2025 | **Revised:** 8 July 2025 | **Accepted:** 15 July 2025

Funding: This work was supported by the Ministry of Education, Culture, Sports, Science and Technology 'Innovation Platform for Society 5.0' (JPMXP0518071489) and the SENSHIN Medical Research Foundation Overseas Study grant.

Keywords: dementia | depression | loneliness | proof of concept | robot | usability

ABSTRACT

Background: Social isolation and loneliness are significant risk factors for poor mental health in older adults, particularly those living alone with cognitive impairment. Socially assistive robots (SARs) offer a promising approach to enhance social connection and emotional well-being in this population. This study aimed to examine the feasibility, acceptability and potential psychosocial impact of a 16-week, home-based use of a conversational SAR in cognitively impaired older adults living alone.

Methods: This was a single-arm, exploratory feasibility study involving home installation of a humanoid conversational robot (RoBoHoN) for 16 weeks. Usability was assessed with the System Usability Scale (SUS). Psychosocial outcomes, including loneliness (UCLA-LS3) and depressive symptoms (GDS-15), were measured pre- and post-intervention. Semi-structured interviews were conducted and analysed using hybrid content analysis to explore user experiences.

Results: A total of 11 participants were enrolled, including individuals with mild cognitive impairment, mild dementia and late-onset psychosis; nine completed the study. The mean age was 81.9 years, with an age range of 71–89 years. The mean SUS score was 81.7, indicating high usability even among cognitively impaired users. Small-to-moderate effect sizes were observed for reductions in loneliness and depressive symptoms (UCLA-LS3: $r = 0.40$; GDS-15: $r = 0.64$), although not statistically significant. Qualitative findings highlighted perceived companionship and increased opportunities to speak, while also noting limitations such as poor conversational depth, delayed responses and technical issues like battery depletion or sensitivity to background noise.

Yuma Nagata and Yuto Satake contributed equally to this study.

Conclusions: The 16-week home-based use of a conversational robot was feasible and well accepted by older adults with cognitive impairment living alone. Although preliminary, findings suggest potential psychosocial benefits of SARs in this population. These results support the rationale for conducting a controlled trial targeting loneliness and depression in cognitively impaired, socially isolated older adults.

1 | Introduction

Social connection is a fundamental human need [1]. Social isolation (the objective lack of social contact) and loneliness (the subjective feeling of being alone) are distinct but related concepts [2], both of which have been associated with increased risks of mortality, physical health problems and mental disorders [3]. These issues are particularly relevant in older people. Longitudinal studies have consistently shown associations between loneliness and depression and anxiety among older people [4–6]. Recent reports estimate that approximately 25% of older people experience social isolation globally [7]. Global recognition of this issue has led to policy responses, including the launch of the WHO Commission on Social Connection in 2023 [8]. These efforts underscore the urgent need to develop accessible interventions to combat social disconnection in later life.

Socially assistive robots (SARs) are one emerging solution that may support social engagement and well-being in older populations. Unlike task-oriented robots, SARs are designed to interact through speech, gesture and presence, creating the potential to reduce loneliness and improve mood through companionship and conversation. Several studies have reported the feasibility and potential psychological benefits of conversational SARs in community-dwelling older adults [9, 10]. However, no studies have examined the psychological effects of conversational SARs use in community-dwelling older people living alone with cognitive impairment. Given prior evidence that older people with mild cognitive impairment (MCI) are particularly vulnerable to loneliness [11–13], this population represents an important potential target for intervention development. To address this, our group has previously explored the potential of a humanoid companion robot, RoBoHoN (Sharp), for use by cognitively impaired older people living alone. In earlier studies, we observed that participants with cognitive decline corresponding to MCI to mild dementia showed positive engagement in communication with the robot for periods of up to 3 months [14] and showed positive attachment to the robot even after a two-month withdrawal period [15]. We have also reported 18-month use of the robot in five participants [16].

This study aimed to examine the feasibility, acceptability, and potential psychosocial impact of a home-based, 16-week conversational-robot (RoBoHoN) intervention in cognitively impaired older people living alone. A mixed-methods approach was employed to explore both quantitative trends (e.g., depressive symptoms and loneliness) and qualitative insights from participants and their caregivers to inform the design of future controlled trials. This study addresses a critical gap in the literature by focusing on cognitively impaired older adults living alone—a population highly susceptible to social isolation—and examining the potential of a conversational robot as a form of home-based psychosocial support.

2 | Methods

2.1 | Study Design

This was a single-arm, observational feasibility study employing a convergent mixed-methods approach [17]. Although the study involved deploying a conversational robot in participants' homes, no randomisation or control group was implemented, and no clinical treatment decisions were altered. The purpose was to explore the acceptability, usability, and psychological impact of the intervention in a real-world setting to inform future controlled trials.

The study was approved as an observational feasibility study by the Osaka University Clinical Research Review Committee (No. 19174). All participants gave written informed consent. This observational study was therefore not registered. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Declaration of Helsinki of 1964, as revised in 2013.

2.2 | Participants

Participants were recruited from among individuals who visited the memory clinics of the Department of Psychiatry, Osaka University Hospital, and Kyowakai Hospital, located in Suita, Osaka, Japan, between March 2021 and August 2021. The eligibility criteria were [1] aged 60 years or older, [2] with mild cognitive decline staged at 0.5 or 1 on the Clinical Dementia Rating (CDR), [3] living alone at home or in private rooms of residence for seniors, and [4] without experience of using the robot. Geriatric psychiatrists conducted diagnoses of dementia and MCI based on international criteria [18, 19]. As a dementia type, Alzheimer's disease (AD) dementia was diagnosed with international criteria [18]. We enrolled patients with MCI, and late-onset psychosis was also diagnosed according to the international criteria [20]. All participants gave informed consent for this study after receiving in-person explanations from Y.S. or Y.N.

Following Whitehead et al., who recommend a minimum of 10–15 participants per group to estimate the outcome standard deviation with approximately 20% precision for a medium standardised effect [21], we set 10 participants as the lower recruitment target and, allowing for an anticipated dropout of up to 20%, aimed to recruit 12 participants for this single-arm feasibility study.

2.3 | Robot

The robot we deployed in this study—RoBoHoN (Sharp Corporation, Osaka, Japan)—is an automated, conversational, humanoid robot of about 20 cm in height and 400 g in weight

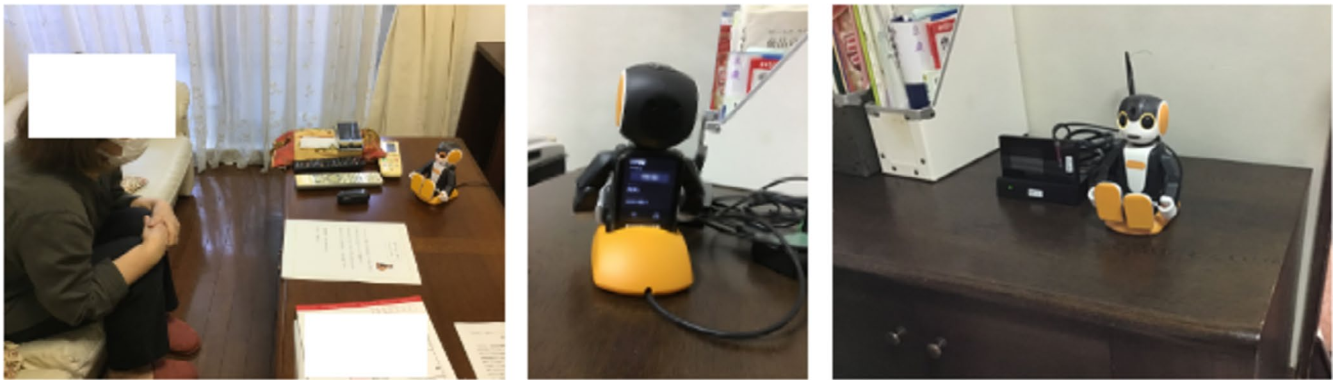

FIGURE 1 | Installation of RoBoHoN in participants' homes. Left: A participant interacting with the robot. Center: The touchscreen on the back of the robot. Right: The Wi-Fi hotspot and the robot. The installation site was selected in consultation with participants and caregivers, aiming to place the robot in a location that would encourage interaction without obstructing daily movement. Although the Wi-Fi hotspot did not need to be positioned immediately next to the robot, it was typically placed in the same room. At the time of installation, printed instructions were provided; however, some caregivers supplemented these with handwritten notes placed near the robot.

(Figure 1). It can perform verbal communication with gestures and entertaining behaviours such as singing. Its conversation is scenario-based. Although it could respond to simple questions (e.g., 'Good morning', 'sing a song' or 'tell me about Osaka'), it could not respond to complicated speech. The robots used for this study were customised to detect conversation logs and make user interaction simpler and faster. The details of the robot are reported in our relevant articles [14, 16], and the setup details are shown in Note S1. An interaction between the robot and a participant is shown in Video S1.

2.4 | Data Collection

Before installation of the robot, we collected the participants' demographic data, including age, gender, education, Mini-Mental State Examination (MMSE) score, and the Lubben Social Network Scale (LSNS-6) assessing social isolation; a lower score (<13 points) is considered to represent an individual who is physically and socially isolated [22]. If the scales had been clinically assessed within the past 3 months, those results were used, and the assessments were not repeated. Those data were retrieved from electronic health records at the recruitment sites.

Assessments of participants' psychological state were conducted twice at their homes: once at baseline/installation and once at removal. The University of California Los Angeles Loneliness Scale, the third version (UCLA-LS3) for loneliness [23, 24], and the 15-item Geriatric Depression Scale (GDS-15) [25] and Hospital Anxiety and Depression Scale Depression Subscale (HADS-D) for depressive mood [26], the State–Trait Anxiety Inventory form Y-2 (STAI Y-2) [27, 28] and HADS Anxiety Subscale (HADS-A) for anxiety [26], the Penn State Worry Questionnaire (PSWQ) for worries [29], the revised Philadelphia Geriatric Centre Morale Scale (R-PGCMS) for subjective happiness [30] and the Short Form-8 Mental Component Summary (SF-8 MCS) for health-related quality of life [31] were used. The participants' general neuropsychiatric symptoms were assessed using the Neuropsychiatric Inventory (NPI)-plus [32, 33]. Although no formal primary outcome was prespecified, the UCLA-LS3 and GDS-15 were of particular interest based on the study's conceptual focus and the earlier literature. The other assessments were selected by the researchers because they were thought to be important potential outcomes and were translated into Japanese.

At the removal, we also administered the System Usability Scale (SUS) [34], our original questionnaire, and a semi-structured interview. We interpreted a score of more than 61 on SUS, indicating that the social robot was accepted by older people in this study based on previous reports [35]. We also administered an original questionnaire consisting of seven items: (1) Appearance, (2) Joy, (3) Comfort, (4) Daily rhythm improvement, (5) Opportunities to speak, (6) Usefulness and (7) Desire to continue use. Each item was rated on a 5-point Likert scale from 1 to 5. While the verbal anchors for 1 and 5 differed slightly across items, all items were designed such that higher scores consistently indicated more favourable opinions. The interview was conducted with the participants and the caregivers to investigate their impression of the strengths and weaknesses of the robot and their suggestions for functions that could be added. The responses were summarised into small sentences and written down on a sheet by the interviewers without audio or visual recording. Field notes were simultaneously recorded in the margins of the paper. The sheet for the original questionnaire and the interview is shown in Figure S1.

All data other than the demographic data were collected by Y.N. and Y.S. at the participants' homes. Some demographic data were retrieved from the electronic health records.

2.5 | Data Analysis

2.5.1 | Quantitative Analysis

The demographic data, SUS scores and our original questionnaire's scores were summarised and described. Spearman's correlation analysis was conducted between SUS and MMSE scores and age because we had speculated that the perceived usability decreases with increased age and decreased cognitive function. Psychological outcomes were assessed at baseline and after 16 weeks using validated self-report measures. Given the small sample size ($n = 9$) and the non-normal distribution of some variables, we used the Wilcoxon signed-rank test as the primary inferential method. Effect sizes for Wilcoxon signed-rank tests were expressed as $r = Z/\sqrt{N}$, where Z is the standardised test statistic and N is the number of paired observations (interpretation: 0.1 small, 0.3 moderate and 0.5 large). For each scale, we report the Hodges–Lehmann median difference and a bias-corrected and accelerated (BCa) 95% confidence interval based on 5000 bootstrap resamples. To account for multiple comparisons among the nine results, p values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure, and corresponding q values were computed. As a robustness check, paired t-tests were performed and summarised in Table S1. All statistical tests were two-tailed with a p value < 0.05, considered statistically significant using R 4.3.3 and relevant statistical packages.

2.5.2 | Qualitative Analysis

We conducted a hybrid content analysis [36, 37], combining deductive coding based on predefined categories with inductive coding to identify emergent subthemes from participants' responses. The coding framework consisted of three hierarchical levels: categories (developed deductively from the research objective), subthemes (identified inductively through close reading of the data) and codes (specific labels assigned to meaningful segments of text). The categories were predefined as 'Strength', 'Weakness' and 'Expected additional functions'.

Two male coders (interviewers)—Yuto Satake (Y.S.; MD, PhD) and Yuma Nagata (Y.N.; OT, PhD)—brought different professional backgrounds to the analysis. Although neither had extensive formal training in qualitative methods at the start of the study, Y.S. had completed a 3-day qualitative research training at University College London and actively participated in a qualitative research self-learning group before the analyses. Both Y.S. and Y.N. had substantial experience in clinical case report writing, particularly in the context of dementia and geriatric psychiatry. The coders were familiar with most participants prior to the interviews, and Y.S. had served as the attending physician for several. Rapport with participants was intentionally cultivated during the pre-intervention home visits.

Y.S. served as the lead coder. The two researchers independently conducted open coding for all interview notes using NVivo 14.24.2 (QSR International, Burlington, MA, USA). Discrepancies in coding were resolved through collaborative discussion, and subthemes were iteratively refined. Inter-coder reliability was not formally quantified, consistent with interpretive qualitative approaches that emphasise reflexive agreement over numerical concordance.

3 | Results

3.1 | Participants

We recruited 11 participants in this study. Two participants dropped out after 3 months of use, and nine completed all assessments. We included data from nine participants completing all assessments in the analyses: two with MCI, six with mild AD dementia and one with late-onset psychosis with a cognitive impairment equivalent to MCI. The demographic data of each participant is shown in Table 1. The mean age and education years (standard deviation: SD) were 81.9 (5.3) and 12.7 (2.8) years, and seven (78%) participants were female. The age range was 71–89 years. The mean MMSE score was 23.6 (3.0). One participant refused to complete the MMSE. The mean LSNS-6 score was 13.2 (1.7), and three participants scored below the cut-off score.

Of the two participants who dropped out, one was an 81-year-old female with MCI who complained that she got frustrated because the robot frequently interrupted her video calls with her family. The other was an 88-year-old female with dementia who stated that she had no time to play with the robot.

3.2 | SUS and Troubles

The mean SUS score (SD) was 81.7 (13.2), ranging from 62.5 to 97.5, which meant all participants showed scores above the threshold score of 61. There were no significant correlations between SUS and MMSE scores (Spearman's $\rho = -0.142$, $p = 0.737$) and age ($\rho = -0.400$, $p = 0.286$). There were a few

TABLE 1 | Participant's demographics ($n = 9$).

ID	Age	Sex	Education	Diagnosis	CDR	MMSE	LSNS-6	UCLA-LS3	GDS-15
P1	71	F	12	Mild AD dementia	0.5	20	18	35	5
P2	77	F	12	MCI	0.5	27	9	41	6
P3	80	F	9	Mild AD dementia	1	27	14	38	4
P4	83	F	9	MCI	0.5	24	21	34	4
P5	83	F	12	Late-onset psychosis	0.5	21	14	37	5
P6	83	F	16	Mild AD dementia	1	22	15	36	6
P7	85	M	16	Mild AD dementia	1	21	15	43	3
P8	86	M	16	Mild AD dementia	1	27	7	51	5
P9	89	F	12	Mild AD dementia	1	NA	6	32	3

Note: The ID numbers correspond to the patient ID shown in the manuscript. F and M mean female and male, respectively. Education is presented as years.
Abbreviations: AD, Alzheimer's disease; CDR, Clinical Dementia Rating; GDS-15, 15-item Geriatric Depression Scale; LSNS-6, Lubben Social Network Scale; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; NA, not applicable; UCLA-LS3, University of California Los Angeles Loneliness Scale, third edition.

phone calls from the participants, and some of them needed our home visits for repairs. All problems were related to the Wi-Fi connection and battery depletion of the robot. There were no harmful events reported by the participants and their caregivers.

3.3 | Original Questionnaire

The mean scores (SD) were 4.8 (0.5) in Appearance, 4.4 (0.5) in Joy, 4.8 (0.4) in Comfort, 3.8 (0.8) in Daily rhythm improvement, 4.6 (0.5) in Opportunities to speak, 4.3 (1.0) in Usefulness and 4.0 (0.9) in Desire to continue use.

3.4 | Psychological Changes

Table 2 presents changes in psychological outcomes as evaluated by Wilcoxon signed-rank tests. No scale reached statistical significance after FDR correction ($q > 0.05$ for all). The two outcomes of primary interest—UCLA-LS3 and GDS-15—showed median reductions of −2 and −1 points, respectively, with effect sizes suggesting moderate to large effects ($r = 0.40$ and 0.64). Their unadjusted p values were 0.23 and 0.05, with FDR-adjusted q-values of 0.89 and 0.48. Paired t-test results were directionally consistent (see Table S1), and Cohen's d for UCLA-LS3 and GDS-15 were 0.42 and 0.80, supporting the potential for clinical relevance.

3.5 | Qualitative Analysis

The interview analyses elucidated the participants and their caregivers' impressions based on their robot use. Each interview was completed within 30 min. Codes, subthemes and respondent counts are summarised in Table S2. In the following texts, P means patient and C means caregivers (e.g., P1 is a patient and C1 is the caregiver of P1).

3.5.1 | Strength

Two subthemes were identified: Companionship and Usefulness. The most frequently mentioned codes Conversational engagement ($n = 7$), Cute looks and actions ($n = 5$), and Friendly greetings ($n = 5$). Participants valued talking with the robot and receiving responses:

P1 said, 'I can talk with him and get some responses from it. It is nice to have reactions'.

P2 added, 'When I leave the house, he says "Have a nice day" and "Welcome back" to me'.

Some participants enjoyed the robot's presence itself, even when interactions were not reciprocal.

P5 reflected: 'I was happy just to listen to his speech. I felt like I was talking to a human being. Even if he didn't reply, I was happy just to have been able to talk to him'.

This aligns with the code Comfort of one-way talk, as well as Uplifting presence.

P6 remarked: 'At first, I didn't feel close to him at all, but he became more like a family member'.

In the Usefulness subtheme, caregivers especially noted more opportunities to speak ($n = 4$):

C2 said, 'My mother had more opportunities to speak. She is talking to him all the time'.

TABLE 2 | Psychological changes analysed using Wilcoxon's signed-rank test.

Scale	Baseline	16 weeks later	*W*	*p*	*q*	*r* (95% CI)
UCLA-LS3	37.00 (6.00)	35.00 (8.00)	33	0.234	0.888	0.397 (−0.408 to 0.873)
GDS-15	5.00 (1.00)	3.00 (2.00)	39	0.053	0.481	0.644 (−0.184 to 0.880)
HADS-D	8.00 (1.00)	8.00 (4.00)	18	0.633	0.888	−0.159 (−0.718 to 0.563)
HADS-A	6.00 (2.00)	6.00 (2.00)	25	0.810	0.888	0.080 (−0.563 to 0.718)
STAI Y-2	43.00 (16.00)	40.00 (7.00)	18.5	0.498	0.888	0.000 (−0.085 to 0.823)
PSWQ	45.00 (11.00)	38.00 (7.00)	31	0.341	0.888	0.317 (−0.441 to 0.880)
R-PGCMS	11.00 (2.00)	11.00 (2.00)	16	0.832	0.888	−0.071 (−0.646 to 0.637)
SF-8 MCS	51.40 (6.94)	52.37 (9.50)	25	0.813	0.888	0.079 (−0.598 to 0.674)
NPI-plus	11.00 (15.00)	6.00 (21.00)	19.5	0.888	0.888	0.000 (−0.306 to 0.823)

Note: The data at baseline and 16 weeks later were presented with median (interquartile range). q = Benjamini–Hochberg false-discovery-rate-adjusted p. r = Wilcoxon effect size ($r = Z/\sqrt{N}$, where N = number of paired observations). 95% confidence intervals derived from 5000 bootstrap resamples.
Abbreviations: GDS-15, 15-item Geriatric Depression Scale; HADS-D, HADS-D/HADS-A, Hospital Anxiety and Depression Scale, Depression/Anxiety subscales; NPI-plus, Neuropsychiatric Inventory-plus; PSWQ, the Penn State Worry Questionnaire; R-PGCMS, the revised Philadelphia Geriatric Centre Morale Scale; SF-8 MCS, Short Form-8 Mental Component Summary; STAI Y-2, State–Trait Anxiety Inventory, form Y-2; UCLA-LS3, University of California Los Angeles Loneliness Scale, third edition.

Other codes included Conversational catalyst and Buffer, with examples such as:

C1: 'My children approached and talked to my mother, triggered by RoBoHoN. He was a good catalyst for conversations'.

C6: 'He acts as a buffer between us'.

3.5.2 | Weakness

Four subthemes were identified: Poor interaction, Poor usability, Poor appearance and Privacy concern.

The most frequent code was Slow or no response ($n=6$), described by P5 as:

'He can give simple responses, but it's difficult for him to carry on a conversation. It would be good if he could respond to you'.

Other problems included Responses to noises and Monotonous speech, mainly noted by caregivers:

C6: 'He has a poor hearing range. He doesn't respond well, especially when a TV is on'.

C3: 'It was annoying when RoBoHoN reacted while my mother was on the phone with her family'.

Under Poor usability, Difficulty in operating ($n=4$) and Difficulty in knowing when to speak ($n=3$) were noted:

C9: 'We needed some support because he stopped responding. It's hard to find the reset button'.

C1: 'It was difficult for my mother to match the timing of her speech with the colour of his eyes'.

Mechanical appearance and small size were noted by two patients, and one caregiver mentioned concerns about recording. Five patients and one caregiver reported no specific weaknesses.

Expected additional functions.

Four subthemes were generated: Daily life support, Family/Caregiver involvement, Personalisation and Better equipment. The most frequent request was Reminder functions ($n=5$):

P4: 'A function that notifies you when you forget to turn off the electricity or gas stove when you go out'.

C9: 'A function that tells you about your schedule, and reminders'.

Other codes included Monitoring and Tailored information:

C8: 'Functions to confirm user's survival and to contact with the person'.

C7: 'I hope he will tell us the news according to our preferences'.

One caregiver suggested remote-controller support. Six patients did not suggest additional functions.

3.5.3 | Other Opinions

Two caregivers commented on emotional attachment forming after 3 months:

C6: 'I gradually started to think he was cute from around the second or third month'.

C9: 'My mother started to be able to talk well with RoBoHoN after about three months of use'.

One participant (P9) gave a striking account of loneliness:

'This robot is the only thing I can talk to now. I can't see anyone else, so no one talks to me. I've ended up all alone...'.

4 | Discussion

This exploratory feasibility study demonstrated that a customised conversational robot can be used continuously for 16 weeks by older people with cognitive impairment who live alone. Retention was high (82%), usability was rated favourably (mean SUS 81.7), and most participants required only minimal technical support. While the sample was small and not powered for hypothesis testing, several outcomes showed small-to-moderate improvements, particularly in depressive symptoms and loneliness. These trends were consistent with participants' qualitative reports of emotional engagement and social connection.

System usability was generally high, with all participants scoring above the standard acceptability threshold. Participants and caregivers reported satisfaction with the robot's appearance, comfort and daily interaction. These findings suggest that a conversational robot can be integrated into the daily routines of cognitively impaired older people with minimal burden. SUS is broadly used to evaluate usability in studies of robots, and the mean score of our study was higher than the reported scores (68.3–80.1) in other reports [38–40]. Our study's dropout rate was comparable to that of other studies [9, 10, 41]. However, two individuals dropped out early. One was frustrated by the robot's interference with phone calls, and the other cited a lack of time, highlighting variability in acceptance. Technical issues were occasionally observed during use, but all were minor enough to be resolved over the phone. Providing an instruction manual and giving an initial orientation may have also been effective. However, it is likely that caregivers provided support to the participants without our knowledge, and this point should be taken into consideration.

Qualitative interviews revealed key strengths of the robot, including its ability to engage in conversational exchanges and offer a comforting presence. The most frequent strength codes were 'Conversational engagement', 'Cute looks and actions' and 'Friendly greetings'. Some participants appreciated the opportunity to talk without expecting responses, reflecting the code ‘Comfort of one-way talk’. Caregivers noted that the robot encouraged participants to speak more, acting as a catalyst for family conversations. Although some intervention studies showed that conversational robots and pet robots did not show apparent differences in psychological support for older people [42, 43], our study suggested that conversational functions may be favourable for older people living alone.

However, several participants and caregivers expressed dissatisfaction with the robot's interactional limitations, such as slow or inappropriate responses. This problem has often been mentioned in previous articles about conversational robots [9, 44]. Many technical issues, like internet connections, voice recognition accuracy and voice generation systems, contribute to this. Above all, we assume the scenario-based conversational system has a significant limitation in making flexible conversations. Future iterations may benefit from incorporating large language models (LLMs) to improve conversational flexibility [45].

Although statistical significance was not achieved after correction for multiple comparisons, the UCLA-LS3 showed a median reduction of 2 points ($r = 0.40$), and the GDS-15 showed a median reduction of 1 point ($r = 0.64$), suggesting potentially meaningful changes in loneliness and depressive symptoms. These findings align with previous studies reporting mood and loneliness improvements with conversational robots [9, 10, 46, 47]. The baseline mean UCLA-LS3 scores were 38.6, lower by as many as 4.6 points than the mean score in older Japanese people [24]—which means they were not so lonely—possibly indicating a ceiling effect. Similarly, none of the participants presented with clinically significant depressive symptoms at baseline, which may have limited the room for improvement. The lack of change in HADS-D scores, despite a moderate effect in GDS-15, may reflect the broader clinical scope of the HADS-D or a floor effect due to low baseline symptoms. The HADS-D includes fewer somatic items and is designed for general hospital settings, which may make it less sensitive to subclinical psychological shifts in older people [48]. In contrast, the GDS is specifically validated for use in older populations and may better detect affective changes in this group [49]. Notably, the PSWQ also showed a numerical decrease, suggesting a possible impact on worry, though further study is needed to interpret this trend.

Although we could not find apparent effects on psychotic symptoms in this study, interestingly, the participant with late-onset psychosis also appeared to benefit from the robot's presence. She remarked that the robot ‘was good not to do unnecessary things’—a sentiment suggesting that the robot functioned as a nonjudgmental listener. This aligns with the qualitative code ‘Comfort of one-way talk’ and supports further exploration of companion robots in populations with paranoid ideation. On the other hand, it is also plausible that some patients with psychosis may become confused or distressed by the robot's responses. For instance, if a robot's dialogue contradicts a delusional belief or misidentification, the patient may experience frustration or exacerbation of symptoms. Such risks could extend to users with memory impairments, especially when robots correct or contradict their narratives. Recent studies have highlighted that LLMs, when applied to companion robots, can inadvertently produce responses that contradict users' beliefs or memories, potentially leading to confusion or distress, particularly in individuals with cognitive impairments or psychosis [50]. Although RoBoHoN used in this study only employed scenario-based conversations and did not produce open-ended responses, these considerations will be important when applying more advanced conversational technologies to populations with psychotic or memory-related symptoms.

The retention rate of 82% and the high usability score support progression to a larger trial, while the small-to-moderate effect sizes observed for loneliness (UCLA-LS3) and depression (GDS-15) suggest that the robot may be a viable intervention for these outcomes in cognitively impaired older people living alone. Using the present variance and effect-size estimates, we calculated sample sizes for a future two-arm trial ($\alpha = 0.05$, $1 - \beta = 0.80$, two-sided t-test, R 4.3.2). For loneliness (SD = 7, $d = 0.40$), approximately 98 participants per group would be required, or 120 participants per group when allowing for a 20% attrition rate. For depressive symptoms ($d = 0.65$), 38 participants per group (48 with 20% attrition) would suffice.

Recruiting isolated older people with cognitive impairment and sustaining in-home robot use for several months will be challenging. A definitive trial will need a two-arm design, randomisation, appropriate blinding of outcome assessment and pre-specified clinically important change thresholds. Further refinement—such as targeting participants with clinically significant loneliness or depression and enhancing conversational responsiveness—will improve both efficiency and potential impact.

This study has several strengths. It targeted a relatively rare and underserved population—older adults with cognitive impairment living alone—and evaluated a conversational robot in a real-world, home-based setting over 16 weeks. The exploratory design included a broad set of validated psychosocial and usability measures, offering a multidimensional view of potential impact. Importantly, participants were recruited from routine clinical outpatient care and had a mean education level of approximately 12 years, reflecting the general older adult population more closely than typical technology-focused samples.

Nonetheless, the study has several limitations that should be acknowledged. First, the small sample size and observational design limit generalisability. Second, the analysed sample likely reflects selection bias, favouring participants more comfortable with technology. Third, although a hybrid coding approach was used, inter-coder reliability was not formally assessed. Fourth, caregiver demographic data were not collected. Fifth, the study period was during the COVID-19 pandemic, which may have influenced the participants' lifestyles. Finally, the 16-week study period may have been insufficient to observe meaningful psychosocial changes.

5 | Conclusion

In summary, our findings suggest that a humanoid conversational robot can be feasibly used over several months by cognitively impaired older people living alone. Quantitative and qualitative data converge in supporting possible psychosocial benefits. These results warrant further investigation in a well-powered external pilot randomised trial with a control group and predefined progression criteria. Future research should include randomised controlled trials with larger sample sizes, explore the use of more advanced conversational technologies such as LLMs, and investigate long-term psychosocial and cognitive outcomes in diverse populations.

Acknowledgements

We show enormous gratitude to Professor Rob Howard, Professor Penny Rapaport and Zuyu Wang at the Division of Psychiatry, University College London, for their valuable input in this manuscript. We also thank the QUDDLE group at the ICLIC, University College London, for their thoughtful suggestions in the qualitative analysis of this study. The authors used ChatGPT (version GPT-4, OpenAI) to improve the language and readability of the manuscript. The authors reviewed and edited all AI-assisted content to ensure accuracy and integrity.

Ethics Statement

This study was approved by the Ethical Committee of Osaka University Hospital.

Consent

Written informed consent for the publication of this report was obtained from the patient.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data sets presented in this article are available upon reasonable request from the corresponding author, Yuto Satake, Email: y.satake@psy.med.osaka-u.ac.jp.

References

1. R. F. Baumeister and M. R. Leary, "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation," *Psychological Bulletin* 117, no. 3 (1995): 497–529.

2. H. O. Taylor, T. K. M. Cudjoe, F. Bu, and M. H. Lim, "The State of Loneliness and Social Isolation Research: Current Knowledge and Future Directions," *BMC Public Health* 23, no. 1 (2023): 1049.

3. J. Holt-Lunstad, "Social Connection as a Critical Factor for Mental and Physical Health: Evidence, Trends, Challenges, and Future Implications," *World Psychiatry* 23, no. 3 (2024): 312–332.

4. S. L. Lee, E. Pearce, O. Ajnakina, et al., "The Association Between Loneliness and Depressive Symptoms Among Adults Aged 50 Years and Older: A 12-Year Population-Based Cohort Study," *Lancet Psychiatry* 8, no. 1 (2021): 48–57.

5. Z. I. Santini, P. E. Jose, E. York Cornwell, et al., "Social Disconnectedness, Perceived Isolation, and Symptoms of Depression and Anxiety Among Older Americans (NSHAP): A Longitudinal Mediation Analysis," *Lancet Public Health* 5, no. 1 (2020): e62–e70.

6. J. Domènech-Abella, J. Mundó, J. M. Haro, and M. Rubio-Valera, "Anxiety, Depression, Loneliness and Social Network in the Elderly: Longitudinal Associations From the Irish Longitudinal Study on Ageing (TILDA)," *Journal of Affective Disorders* 246 (2019): 82–88.

7. R. H. Teo, W. H. Cheng, L. J. Cheng, Y. Lau, and S. T. Lau, "Global Prevalence of Social Isolation Among Community-Dwelling Older Adults: A Systematic Review and Meta-Analysis," *Archives of Gerontology and Geriatrics* 107 (2023): 104904.

8. World Health Organization, "WHO Commission on Social Connection [Internet]," 2024, https://www.who.int/groups/commission-on-social-connection.

9. O. E. Lee, K. O. Nah, E. M. Kim, N. G. Choi, and D. H. Park, "Exploring the Use of Socially Assistive Robots Among Socially Isolated Korean American Older Adults," *Journal of Applied Gerontology* 43, no. 9 (2024): 1295–1304.

10. O. E. K. Lee, I. Nam, Y. Chon, A. Park, and N. Choi, "Socially Assistive Humanoid Robots: Effects on Depression and Health-Related Quality of Life Among Low-Income, Socially Isolated Older Adults in South Korea," *Journal of Applied Gerontology* 42, no. 3 (2023): 367–375.

11. L. Boss, D. H. Kang, and S. Branson, "Loneliness and Cognitive Function in the Older Adult: A Systematic Review," *International Psychogeriatrics* 27, no. 4 (2015): 541–553.

12. J. Yu, C. L. M. Lam, and T. M. C. Lee, "Perceived Loneliness Among Older Adults With Mild Cognitive Impairment," *International Psychogeriatrics* 28, no. 10 (2016): 1681–1685.

13. C. R. Victor, I. Rippon, S. M. Nelis, et al., "Prevalence and Determinants of Loneliness in People Living With Dementia: Findings From the IDEAL Programme," *International Journal of Geriatric Psychiatry* 35, no. 8 (2020): 851–858.

14. D. Figueroa, R. Yamazaki, S. Nishio, et al., "Social Robot for Older Adults With Cognitive Decline: A Preliminary Trial," *Frontiers in Robotics and AI* 10 (2023): 1213705.

15. R. Yamazaki, S. Nishio, Y. Nagata, et al., "Long-Term Effect of the Absence of a Companion Robot on Older Adults: A Preliminary Pilot Study," *Frontiers in Computer Science* 5 (2023): 1129506.

16. Y. Nagata, Y. Satake, R. Yamazaki, S. Nishio, H. Ishiguro, and M. Ikeda, "Use of a Companion Robot at Home by Older People With Cognitive Decline and Psychosis: A Case Series With 18-Month Follow-Up [Internet]," 2024, https://www.researchsquare.com/article/rs-5676562/v1.

17. J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. (SAGE, 2018), 492.

18. G. M. McKhann, D. S. Knopman, H. Chertkow, et al., "The Diagnosis of Dementia due to Alzheimer's Disease: Recommendations From the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease," *Alzheimer's & Dementia* 7, no. 3 (2011): 263–269.

19. M. S. Albert, S. T. DeKosky, D. Dickson, et al., "The Diagnosis of Mild Cognitive Impairment due to Alzheimer's Disease: Recommendations From the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease," *Alzheimer's & Dementia* 7, no. 3 (2011): 270–279.

20. R. Howard, P. V. Rabins, M. V. Seeman, and D. V. Jeste, "Late-Onset Schizophrenia and Very-Late-Onset Schizophrenia-Like Psychosis: An International Consensus," *American Journal of Psychiatry* 157, no. 2 (2000): 172–178.

21. A. L. Whitehead, S. A. Julious, C. L. Cooper, and M. J. Campbell, "Estimating the Sample Size for a Pilot Randomised Trial to Minimise the Overall Trial Sample Size for the External Pilot and Main Trial for a Continuous Outcome Variable," *Statistical Methods in Medical Research* 25, no. 3 (2016): 1057–1073.

22. J. Lubben, E. Blozik, G. Gillmann, et al., "Performance of an Abbreviated Version of the Lubben Social Network Scale Among Three European Community-Dwelling Older Adult Populations," *Gerontologist* 46, no. 4 (2006): 503–513.

23. D. W. Russell, "UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure," *Journal of Personality Assessment* 66, no. 1 (1996): 20–40.

24. Y. Masuda, E. Tadaka, and Y. Dai, "Reliability and Validity of the Japanese Version of the UCLA Loneliness Scale Version 3 Among the Older Population," *Japanese Academy of Community Health Nursing* 15 (2012): 25–32.

25. K. Sugishita, M. Sugishita, I. Hemmi, T. Asada, and T. Tanigawa, "A Validity and Reliability Study of the Japanese Version of the Geriatric Depression Scale 15 (GDS-15-J)," *Clinical Gerontologist* 40, no. 4 (2017): 233–240.

26. A. S. Zigmond and R. P. Snaith, "The Hospital Anxiety and Depression Scale," *Acta Psychiatrica Scandinavica* 67, no. 6 (1983): 361–370.

27. N. Iwata, N. Mishima, T. Shimizu, et al., "The Japanese Adaptation of the STAI Form Y in Japanese Working Adults. The Presence or Absence of Anxiety," *Industrial Health* 36, no. 1 (1998): 8–13.

28. C. Spielberger, R. Gorsuch, R. Lushene, P. Vagg, and G. Jacobs, *Manual for the State–Trait Anxiety Inventory (Form Y)* (Consulting Psychologists Press, 1983).

29. T. J. Meyer, M. L. Miller, R. L. Metzger, and T. D. Borkovec, "Development and Validation of the Penn State Worry Questionnaire," *Behaviour Research and Therapy* 28, no. 6 (1990): 487–495.

30. M. P. Lawton, "The Philadelphia Geriatric Center Morale Scale: A Revision," *Journal of Gerontology* 30, no. 1 (1975): 85–89.

31. S. Fukuhara and Y. Suzukamo, *Manual of the SF-8 Japanese Version* (Qualitest, 2004).

32. J. L. Cummings, M. Mega, K. Gray, S. Rosenberg-Thompson, D. A. Carusi, and J. Gornbein, "The Neuropsychiatric Inventory: Comprehensive Assessment of Psychopathology in Dementia," *Neurology* 44, no. 12 (1994): 2308.

33. E. Mori, M. Ikeda, K. Kosaka, and the Donepezil-DLB Study Investigators, "Donepezil for Dementia With Lewy Bodies: A Randomized, Placebo-Controlled Trial," *Annals of Neurology* 72, no. 1 (2012): 41–52.

34. J. Brooke, "SUS—A Quick and Dirty Usability Scale," in *Usability Evaluation in Industry*, ed. J. Brooke (Taylor & Francis, 1996), 189–194.

35. A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale," *International Journal of Human-Computer Interaction* 24, no. 6 (2008): 574–594.

36. J. Fereday and E. Muir-Cochrane, "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development," *International Journal of Qualitative Methods* 5, no. 1 (2006): 80–92.

37. K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 4th ed. (SAGE, 2019), 453.

38. R. Boumans, F. van Meulen, K. Hindriks, M. Neerincx, and R. M. Olde, "A Feasibility Study of a Social Robot Collecting Patient Reported Outcome Measurements From Older Adults," *International Journal of Social Robotics* 12, no. 1 (2020): 259–266.

39. F. Liang, Z. Su, W. Sheng, A. Bishop, and B. Carlson, "Medication Adherence Management for In-Home Geriatric Care With a Companion Robot and a Wearable Device," *Smart Health* 30 (2023): 100434.

40. A. M. Asl, M. M. Ulate, M. F. Martin, and H. van der Roest, "Methodologies Used to Study the Feasibility, Usability, Efficacy, and Effectiveness of Social Robots for Elderly Adults: Scoping Review," *Journal of Medical Internet Research* 24, no. 8 (2022): e37434.

41. M. Tanaka, A. Ishii, E. Yamano, et al., "Effect of a Human-Type Communication Robot on Cognitive Function in Elderly Women Living Alone," *Medical Science Monitor* 18, no. 9 (2012): CR550–CR557.

42. H. Robinson, B. MacDonald, N. Kerse, and E. Broadbent, "The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial," *Journal of the American Medical Directors Association* 14, no. 9 (2013): 661–667.

43. M. Valentí Soler, L. Agüera-Ortiz, J. Olazarán Rodríguez, et al., "Social Robots in Advanced Dementia," *Frontiers in Aging Neuroscience* 7 (2015): 00133, http://journal.frontiersin.org/Article/10.3389/fnagi.2015.00133/abstract.

44. B. Irfan, S. Kuoppamäki, and G. Skantze, "Recommendations for Designing Conversational Companion Robots With Older Adults Through Foundation Models," *Frontiers in Robotics and AI* 11 (2024): 1363713, https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2024.1363713/full.

45. Y. Satake, D. Ishimaru, K. Katsuki, et al., "A Week With a Conversational Large Language Model Companion Robot," *American Journal of Geriatric Psychiatry* 33 (2025): S1064748125001071.

46. J. Lim, "Effects of a Cognitive-Based Intervention Program Using Social Robot PIO on Cognitive Function, Depression, Loneliness, and Quality of Life of Older Adults Living Alone," *Frontiers in Public Health* 11 (2023): 1097485.

47. N. Fields, L. Xu, J. Greer, and E. Murphy, "Shall I Compare Thee…to a Robot? An Exploratory Pilot Study Using Participatory Arts and Social Robotics to Improve Psychological Well-Being in Later Life," *Aging & Mental Health* 25, no. 3 (2021): 575–584.

48. I. Bjelland, A. A. Dahl, T. T. Haug, and D. Neckelmann, "The Validity of the Hospital Anxiety and Depression Scale: An Updated Literature Review," *Journal of Psychosomatic Research* 52, no. 2 (2002): 69–77.

49. J. A. Yesavage, T. L. Brink, T. L. Rose, et al., "Development and Validation of a Geriatric Depression Screening Scale: A Preliminary Report," *Journal of Psychiatric Research* 17, no. 1 (1982): 37–49.

50. B. Irfan, S. Kuoppamäki, A. Hosseini, and G. Skantze, "Between Reality and Delusion: Challenges of Applying Large Language Models to Companion Robots for Open-Domain Dialogues With Older Adults," *Autonomous Robots* 49, no. 1 (2025): 9.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1.** The sheet for the original questionnaire and the interview. **Note S1.** Robot setup. **Table S1.** Psychological changes analysed using paired *t*-test. **Table S2.** Codes, subthemes and categories of qualitative data. **Video S1.** The interaction between the robot and a participant.