Original Paper

"Digital Clinicians" Performing Obesity Medication Self-Injection Education: Feasibility Randomized Controlled Trial

Sean Coleman[1,2], MBBChBAO, MSc; Caitríona Lynch[1], BSc, MSc; Hemendra Worlikar[2], BEng, MEng; Emily Kelly[2]; Kate Loveys[3], PhD; Andrew J Simpkin[4], PhD; Jane C Walsh[5], PhD; Elizabeth Broadbent[3], PhD; Francis M Finucane[1,2], MD; Derek O' Keeffe[1,2], MD, MBA, MEng, PhD

[1]Department of Diabetes, Endocrinology, and Metabolism, University Hospital Galway, Galway, Ireland
[2]School of Medicine, University of Galway, Galway, Ireland
[3]School of Psychology, University of Auckland, Auckland, New Zealand
[4]School of Mathematics and Statistical Science, University of Galway, Galway, Ireland
[5]School of Psychology, University of Galway, Galway, Ireland

Corresponding Author:

Sean Coleman, MBBChBAO, MSc
Department of Diabetes, Endocrinology, and Metabolism
University Hospital Galway
Newcastle Rd
Galway H91 YR71
Ireland
Phone: 353 851604880
Email: seancoleman@live.ie

Abstract

Background: Artificial intelligence (AI) chatbots have shown competency in a range of areas, including clinical note taking, diagnosis, research, and emotional support. An obesity epidemic, alongside a growth in novel injectable pharmacological solutions, has put a strain on limited resources.

Objective: This study aimed to investigate the use of a chatbot integrated with a digital avatar to create a "digital clinician." This was used to provide mandatory patient education for those beginning semaglutide once-weekly self-administered injections for the treatment of overweight and obesity at a national center.

Methods: A "digital clinician" with facial and vocal recognition technology was generated with a bespoke 10- to 15-minute clinician-validated tutorial. A feasibility randomized controlled noninferiority trial compared knowledge test scores, self-efficacy, consultation satisfaction, and trust levels between those using the AI-powered clinician avatar onsite and those receiving conventional semaglutide education from nursing staff. Attitudes were recorded immediately after the intervention and again at 2 weeks after the education session.

Results: A total of 43 participants were recruited, 27 to the intervention group and 16 to the control group. Patients in the "digital clinician" group were significantly more knowledgeable postconsultation (median 10, IQR 10-11 vs median 8, IQR 7-9.3; *P*<.001). Patients in the control group were more satisfied with their consultation (median 7, IQR 6-7 vs median 7, IQR 7-7; *P*<.001) and had more trust in their education provider (median 7, IQR 4.8-7 vs median 7, IQR 7-7; *P*<.001). There was no significant difference in reported levels of self-efficacy (*P*=.57). 81% (22/27) participants in the intervention group said they would use the resource in their own time.

Conclusions: Bespoke AI chatbots integrated with digital avatars to create a "digital clinician" may perform health care education in a clinical environment. They can ensure higher levels of knowledge transfer yet are not as trusted as their human counterparts. "Digital clinicians" may have the potential to aid the redistribution of resources, alleviating pressure on bariatric services and health care systems, the extent to which remains to be determined in future studies.

Trial Registration: ISRCTN ISRCTN12382879; https://www.isrctn.com/ISRCTN12382879

JMIR Diabetes 2025;10:e63503; doi: 10.2196/63503

Keywords: automation; automate; machine-human interface; clinical education; digital clinician; virtual human; trust; chatGPT; chatbots; machine learning; ML; artificial intelligence; AI; large language models; natural language processing; NLP; deep learning; randomized controlled trials; RCTs; feasibility studies; obesity; medication

Introduction

Obesity, an "abnormal or excessive accumulation of fat that poses a health risk" [1], is a primary contributor to global health challenges [2]. Reports suggest its involvement in up to 80% of cases of type 2 diabetes and 43% of cardiovascular incidents [3] while significantly contributing to depression and anxiety [4].

Recent breakthroughs in bariatric medicine have elevated the role of injectable therapy, namely glucagon-like peptide-1 agonist agents such as semaglutide. New injectable agents are showing weight loss surpassing 20% and 25% [5]. With over half of adults in the World Health Organization (WHO) European Region [2] potentially eligible, use is hindered by scalability of clinical services and supply. Generative AI has been identified to have the potential to offset the clinical and administrative demands associated with the management of patients on these medication types [6]

Patient education is a critical component of the clinical care pathway and a prerequisite at many clinics for the prescription of pharmacotherapy. Even in resource-rich countries, the necessary services are not always available. Studies have predicted the United States needs careful restructuring of health service expenditure to meet demand for the costs of overweight and obesity services [7].

Health care systems are underresourced and understaffed, particularly in rural areas [8]. This impacts productivity, sustainability [9], and the health of both patient and health care professionals [10]. Health care worker shortages, propelled by aging populations [11] and the COVID-19 pandemic [12], are exacerbated in areas such as the Global South, increasing health care inequities [13]. Tasks that health care workers perform are often repetitive and administrative. Redistribution of such tasks has been shown to potentially improve health outcomes such as blood pressure, HbA_{1c}, and mental health [14].

Artificial Intelligence (AI) chatbots have the potential to be used for a variety of medical tasks, including note taking [15] and personalized medicine [16]. Anonymized AI chatbots have been judged to provide better, more concise, empathetic answers to general health queries than verified physicians [17]. ChatGPT-3 is known to be accurate with common chief complaints [18] and GPT-4 recently outscored 99.98% of simulated human readers when diagnosing complex clinical cases [19]. Automated personalized messaging systems are being researched to enhance behavioral change in a hope to enhance health outcomes [20].

However, there are concerns about trust, usability, and efficacy [21]. Hallucinating models can spread misinformation and private medical data may be misused [16]. Trust is fundamental to the physician-patient relationship shown to affect health behaviors, compliance, and quality of life [22]. Willingness to use supportive technologies has been shown to be influenced by complex factors, such as perceived usefulness, health threat, and resistance to change [23]. Chatbots are text-based and have rarely been integrated with physical form, for example, avatars or "virtual humans." Adding form to a faceless chatbot to create a "digital clinician" may increase trust, engagement, and usability [24].

Chatbots are text-based and have rarely been integrated with physical form, for example, avatars or "virtual humans." Adding form to a faceless chatbot to create a "digital clinician" may increase trust, engagement, and usability [24].

Automation may offer benefits in standardization, efficiency, effectiveness, cost, confidentiality, and access. Social desirability response bias is associated with higher levels of treatment nonadherence [25] and reduces the accuracy of clinical history taking [26] in human-human interactions. In educational settings, certain interactions, such as quizzing, enhance information retention but may be more socially appropriate from a "digital clinician" than from a health care professional.

Virtual humans have been shown to be efficacious in nonclinical patient-facing scenarios [27]. There has been a paradigm shift since the advent of ChatGPT, bringing automated communicators into a new light, mandating research focus. The aim of this study is to investigate the use of a medically approved task-specific "digital clinician" to provide patient education in a clinical environment, comparing it with a human counterpart.

Methods

Research Design

All patients at Galway University Hospitals must attend an education session with a Clinical Nurse Specialist before starting semaglutide injectable therapy. This session covers basic semaglutide pharmacotherapy and the safe self-administration of injections using a semaglutide pen. During the study period, allocated, eligible participants received their mandatory education from a "digital clinician" with human oversight. The control group received current standard-of-care, human, nurse-led education.

Once a clinical decision was made that treatment of overweight or obesity with semaglutide injections would be commenced, eligibility was assessed. Overweight and obesity were classified as a BMI greater than 25 kg/m^2. Eligibility criteria included being aged 18 years or older, without an intellectual or physical disability, which would interfere with a participant's ability to self-administer semaglutide injections. All eligible patients were invited to participate in the study. Given the nature of the intervention ("digital clinician"), blinding was not feasible.

Eligibility criteria included patients aged older than 18 years, without an intellectual or physical disability, which would interfere with a participant's ability to self-administer semaglutide injections. All eligible patients were invited to participate in the study. Given the nature of the intervention ("digital clinician"), blinding was not feasible.

The "Digital Clinician"

To inform design, clinicians were observed educating patients. Their mannerisms, behavior, and conversations were recorded. Transcripts of educational sessions given by the local obesity nurse specialist were generated.

A 10- to 15-minute educational script was generated with multiple conversation streams. Information on medication name, dosages, mechanism of action, pen preparation, administration, side effects, and storage was included. User questioning was used to maximize engagement and knowledge retention.

Akin to natural clinical education, the "digital clinician" led the primary section of the tutorial, offering information and asking the patient questions throughout to test retention and assess understanding. The question types used by the "digital clinician" varied. Some were open, while others were multiple choice, true or false, yes or no, or numerical. After the clinician-led portion of the tutorial had ended, patients were invited to type or voice any question they wished. The option to see and hear the answers to some frequently asked questions was also offered by way of onscreen prompts.

Unclear or ambiguous responses were identified by the digital clinician as such, and this was communicated to the user who was then kindly asked to repeat themselves. If not understanding the user twice in a row, the "digital clinician" would simply offer the correct answer and continue. This prevented the possibility of an infinite loop. When offered the chance to ask general queries, the user may query infinitely if they wish, although a comprehensive list of frequently asked question prompts on screen, and an onscreen button to end the tutorial, were in place to reduce this need.

A careful selection of trigger words and combinations of "IF," "AND," "OR," and "NOT" statements was used to design the conversational flow. Rigorous internal testing, including over 100 iterations of the conversational flow, was tested by the research team to ensure a fluent, cohesive, usable digital clinician with little risk of misinterpreting the user's intent.

This "digital clinician" could be accessed with native web browsers on the participants' phone, tablet, or laptop. In this case, the participants accessed the session under supervision on an iPad (Apple) device.

Two multiscene videos were recorded by the team and integrated into the tutorial using screen-in-screen projection. The videos offered human views and instructions of what preparing and using the injection pen entailed. The videos, while being demonstrated by members of the research team, were narrated by the "digital clinician."

The software IBM Watson Assistant was used to generate conversation streams. Microsoft ClipChamp Video Editor was used to edit videos, then uploaded to Imgur, accessed via source code written on IBM Watson. The IBM Watson conversation code file was integrated with an animated avatar on the SoulMachines Creator Website. The "digital clinician" uses natural language processing and an evolving bank of AI-driven animation responses that monitor vocal tone and facial expression to regulate behavior. The "digital clinician" has a programmed personality, name, voice, and identity and is being used in a role traditionally considered to require human intelligence.

Logos of affiliated educational and health care institutions were included on the user interface screen to increase trust. Cinematic screening offered different angles of the avatar's face to make for a more interesting experience. Multiple avatars of different genders, races, and names were used to promote inclusivity. All possible conversation streams had been verified against official pharmaceutical literature. The individual scripts were not recorded, but each participant-AI interaction was supervised to monitor for technical issues or hallucinations. Multiple avatars of different genders, races, and names were used to promote inclusivity. All possible conversation streams had been verified against official pharmaceutical literature. The individual scripts were not recorded, but each participant-AI interaction was supervised to monitor for technical issues or hallucinations. A short video of the "digital clinician" in use is included in Multimedia Appendix 1

Randomization: Minimization

Minimization is a method used in clinical trials and experimental research to allocate participants [4,28]. Participants were allocated in order of enrollment to the study arm that minimized the difference across 4 selected numerical baseline variables, namely, age, BMI, pretutorial knowledge of semaglutide score, and pretutorial injection self-efficacy score. Order of enrollment in the study was determined by check-in time at the clinic. Allocation by minimization is deterministic and hence precludes traditional concealment.

The first participant was assigned at random using a coin toss. The second participant was subsequently assigned to the alternative group. Allocation was verified using Microsoft Excel by an independent assessor with access to study data in real-time at a separate location via a secure web-based suppository. The assessor then communicated the allocation to the research team on site.

Allocation was verified using Microsoft Excel by an independent assessor with access to study data in real-time at a separate location via a secure online suppository. The assessor then communicated the allocation to the research team on site.

Self-Efficacy

Self-efficacy was measured using the 2-part validated Self Injection Assessment Questionnaire (SIAQ) [29]. The preinjection SIAQ was used as baseline data. The postinjection SIAQ was scored across 5 domains considering

2-week postintervention feelings about injections, self-image, self-confidence, pain and skin reactions, ease of use, and satisfaction. The SIAQ has been shown to be a valid, robust tool with sufficient validity, reliability, consistency, and sensitivity. Cronbach α and the test-retest coefficient were >0.70 for all domains [30].

Knowledge Attainment

Unvalidated knowledge assessments were designed specifically for the purpose of this research. The pre-educational assessment tool consisted of 4 questions. The posteducational assessment tool contained the same initial 4 questions and a further 8 questions. Questions were multiple choice and covered topics, such as drug name, drug class, mechanism of action, side effects, injection technique, and storage requirements.

Consultation Satisfaction

The validated Patients' Overall Satisfaction with Primary Care Physicians Scale (CPSS) [31] has criterion-related validity coefficients mostly in the 0.80s and 0.90s when considering empathy, physician recommendation, and general satisfaction. Cronbach coefficient α for the patient satisfaction scale is 0.98. The 7-point Likert scale was used in the study.

Trust

The "Trust between People and Automation Scale" (TPA) has undergone validation [32] and been in use for more than 20 years. It was adapted and integrated with the consultation satisfaction measure for this study.

Technology Usability and Future Use

The validated Technology Usability Questionnaire (TUQ) [33] was adapted. The questionnaire has high reliability with Cronbach Coefficient α more than 0.8 across all domains, and high validity expressed through its multiple native questionnaires [34]. A shortened version including questions from 4 of 5 domains was used. The domains were "Ease of use and Learnability," "Interface Quality," "Interaction Quality," and "Satisfaction and future use." Open-ended questions were used to record general themes and attitudes toward the "digital clinician." The outcome measure tools are included as Multimedia Appendix 2.

Sample Size Calculations

When calculating sample size, a similar study using the validated SIAQ to measure self-efficacy was used [29]. It showed a mean 7.09 (SD of 1.5), on a 10-point Likert scale. A significant difference of two-thirds the SD was proposed to boundary noninferiority, with 80% power and α=.05; this would require 28 participants to be recruited into each arm of the study.

Ethical Considerations

A randomized controlled noninferiority trial of "digital clinician"-led patient education was designed with ethical approval from Galway University Hospitals, Clinical Research Ethics Committee (Ref: CA 2920). The trial and trial protocol were guided by SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials involving Artificial Intelligence) [35] and CONSORT-AI (Consolidated Standards of Reporting Trials- Artificial Intelligence) [34] extensions. Further information on the study protocol can be found in Multimedia Appendix 3. It was ensured that all patients were offered at least the current standard of care. After data collection, all patients were seen by the clinical nurse specialist. The nurse assessed the patients who had received safety-net instructions before being discharged with a medication information leaflet. Contact with the nurse minimized disruption of the patient-clinician relationship. This established human oversight and accountability. Informed written consent was obtained from all participants before enrollment and randomization, after provision of a patient information leaflet and discussion with the research team. All patient data was anonymized. No patient identifiable information was included in data analysis and confidentiality was maintained. Data was stored on password-protected encrypted devices. No compensation was offered to participants.

Statistical Analysis and Missing Data

Data were analyzed using Jamovi version 2.4.14 developed by Love, Droppman, and Selker [36] and R Studio, developed by Posit PBC [37]. All Likert scale data were appropriately aggregated in treatment and control arms and recalibrated to scales from 1 to 10 for self-efficacy, and 1-7 for trust, satisfaction, and usability.

The Mann-Whitney U test was used to assess differences in primary and secondary outcome variables. This nonparametric test is robust to the distribution of the outcome data. The Mann-Whitney *U* test compares the ranks of all the data points in 2 groups. *P* values below .05 were deemed statistically significant. Conventional content analysis was used to interpret qualitative data. Missing outcome data were omitted from analysis. Observed outcome data were analyzed as if representative of the entire cohort.

Knowledge, satisfaction, trust, and usability were measured immediately posttutorial. Subsequent loss to follow-up at 2 weeks did not affect the handling of this data.

Those lost to follow-up were unavailable for self-efficacy analysis. The profile of those with complete data and those with missing data were explored in Multimedia Appendix 4.

Results

Baseline Characteristics

During the 7-week study period in the bariatric clinic, 53 patients were prescribed semaglutide self-administered injections for the treatment of overweight or obesity (see Figure 1). Of this cohort, 43 agreed to participate in the study. Of those enrolled, 100% (43/43) completed the pretutorial questionnaire, their assigned education session, and their posttutorial questionnaire. Only 41.9% (18) participants

completed the 2-week follow-up self-efficacy questionnaire (Table 1).

Figure 1. A flowchart of enrollment, allocation, and attrition details.

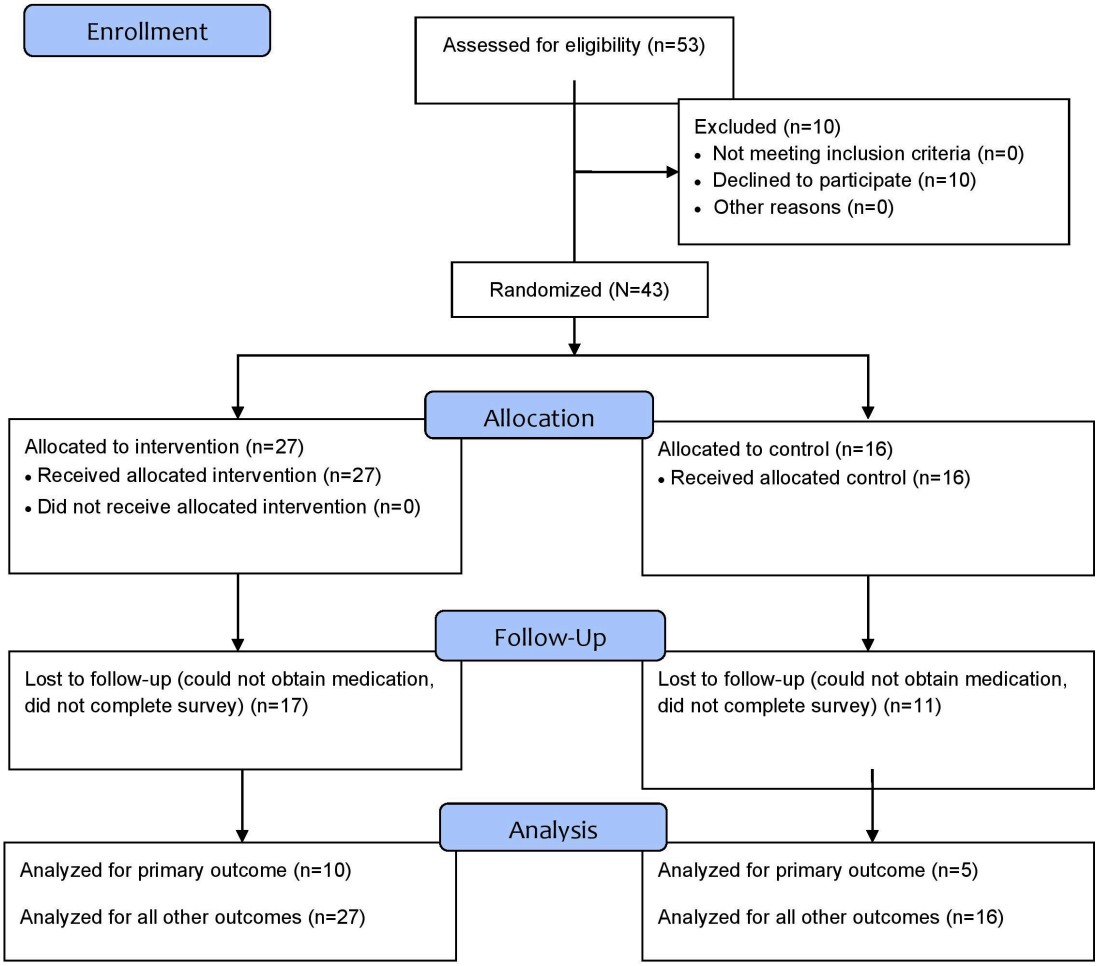


Table 1. A summary of participants' baseline characteristics.

Characteristic	Total (n=43)	Digital clinician (n=27)	Control (n=16)	*P* value
Age, mean (SD)	47.41 (13)	46.4 (13)	49.2 (13)	—
BMI, mean (SD)	43.6 (8)	42.9 (6.8)	44.8 (10)	—
Sex, n (%)				
Male	9 (21)	3 (11)	6 (38)	—
Female	34 (79)	24 (89)	10 (63)	—
Ethnicity, n (%)				
Irish	30 (70)	19 (70)	11 (69)	—
Other	4 (9)	3 (11)	1 (6)	—
Missing	9 (21)	5 (19)	4 (25)	—
Education level, n (%)				
None	1 (2)	1 (4)	0 (0)	—
Primary	2 (5)	1 (4)	1 (6)	—
Secondary	19 (44)	12 (44)	7 (44)	—
Third level	21 (49)	13 (48)	8 (50)	—
Pretutorial Knowledge, median (IQR)	2 (1-2)	2 (1-3)	2 (1-2)	.41
Pretutorial Self Efficacy, median (IQR)	27 (31-33)	32 (27-33)	31 (27.8-33.3)	1.00

Self Efficacy

Both groups had a median self-efficacy level of 10/10, with the "digital clinician" group having a marginally lower first quartile of 8, compared with the control group (8.2; Figure 2). This difference was not statistically significant (Table 2; *P*=.52).

Responses recording "Pain and Skin reactions" from the control group had a favorable, nonsignificantly higher mean rank.

Figure 2. Self-Efficacy scores across control and digital clinician groups. SIAQ: Self-Injection Assessment Questionnaire.

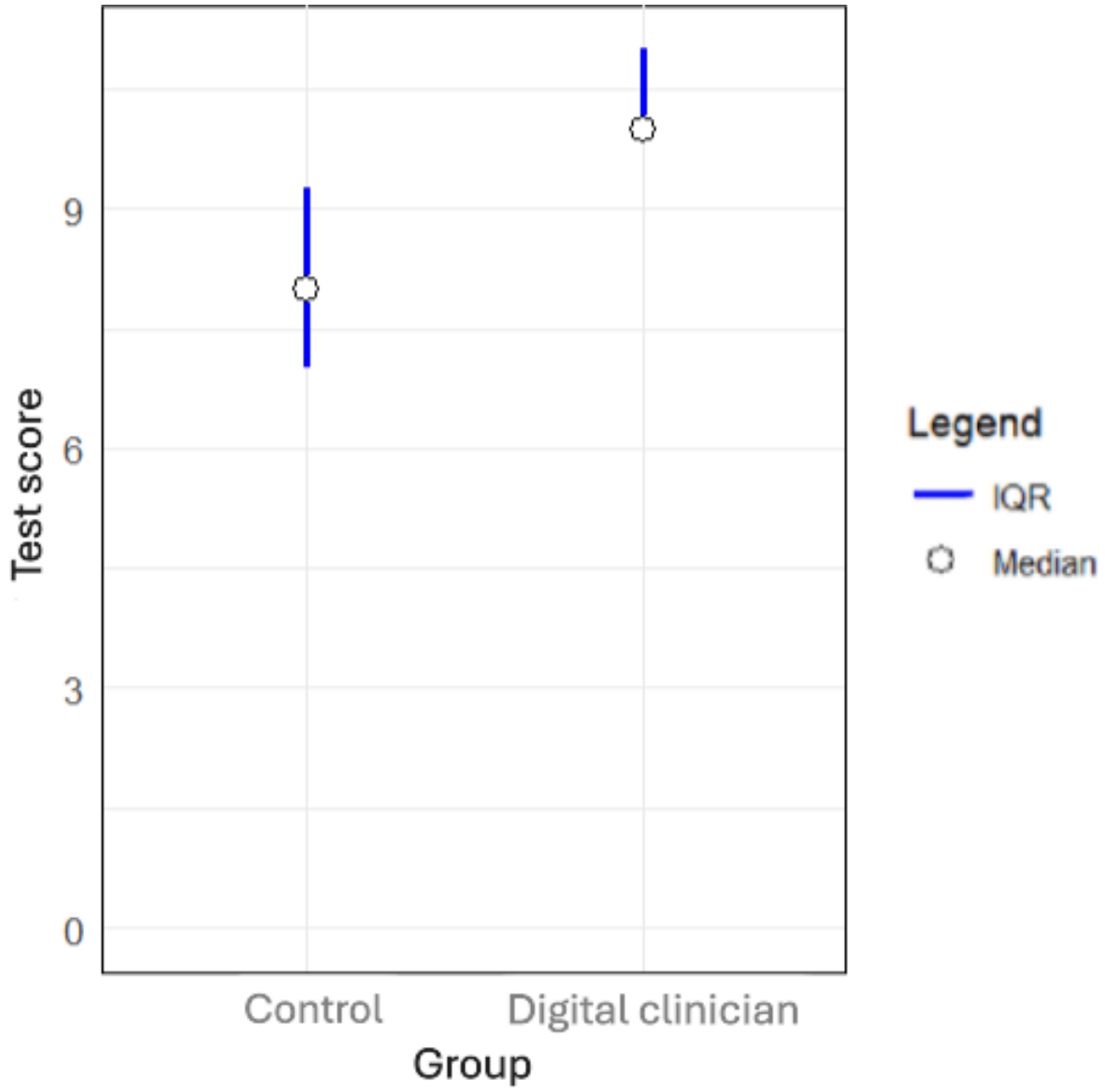


Table 2. Self-injection self-efficacy assessment results at 2 weeks post intervention.

Domain	Digital clinician (n=10), median (IQR)	Control (n=5), median (IQR)	*P* value
Feelings about injections	5.0 (5.0-5.0)	5.0 (4.0-5.0)	.08
Self-image	5.0 (4.0-5.0)	5.0 (5.0-5.0)	.66
Self-confidence	4.5 (4.0-5.0)	5.0 (4.0-5.0)	.45
Pain and skin reactions	5.0 (5.0-5.0)	5.0 (5.0-5.0)	.08
Ease of use	6.0 (5.0-6.0)	5.0 (5.0-6.0)	.22
Satisfaction	5.0 (4.0-5.0)	5.0 (4.0-5.0)	.61
Overall self-efficacy	10.0 (8.0-10.0)	10.0 (8.2-0.0)	.52

Knowledge (Test Score)

Overall, the median test score of participants postconsultation in the "digital clinician" group was 10/12 (84%; IQR 10-11), while the median score in the control group was 8/12 (69%; IQR 7-9; P<.001; Figure 3, Table 3).

Figure 3. Knowledge scores across control and digital clinician groups. SIAQ: Self Injection Assessment Questionnaire.

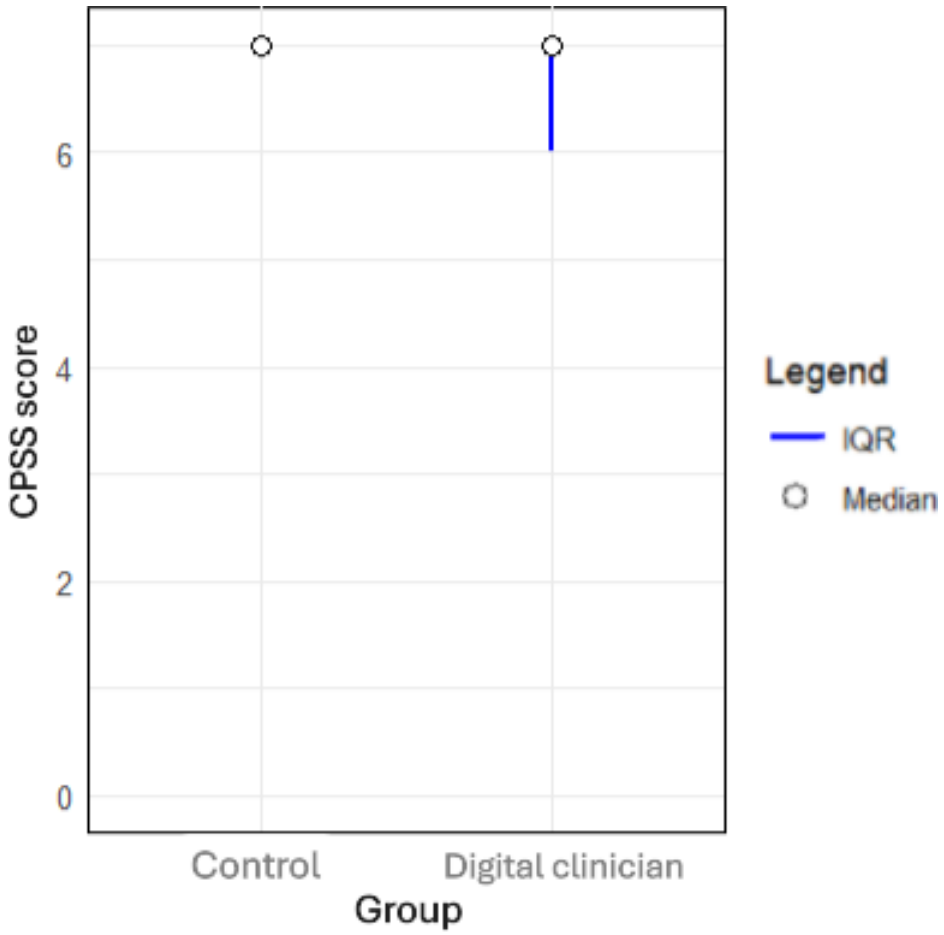


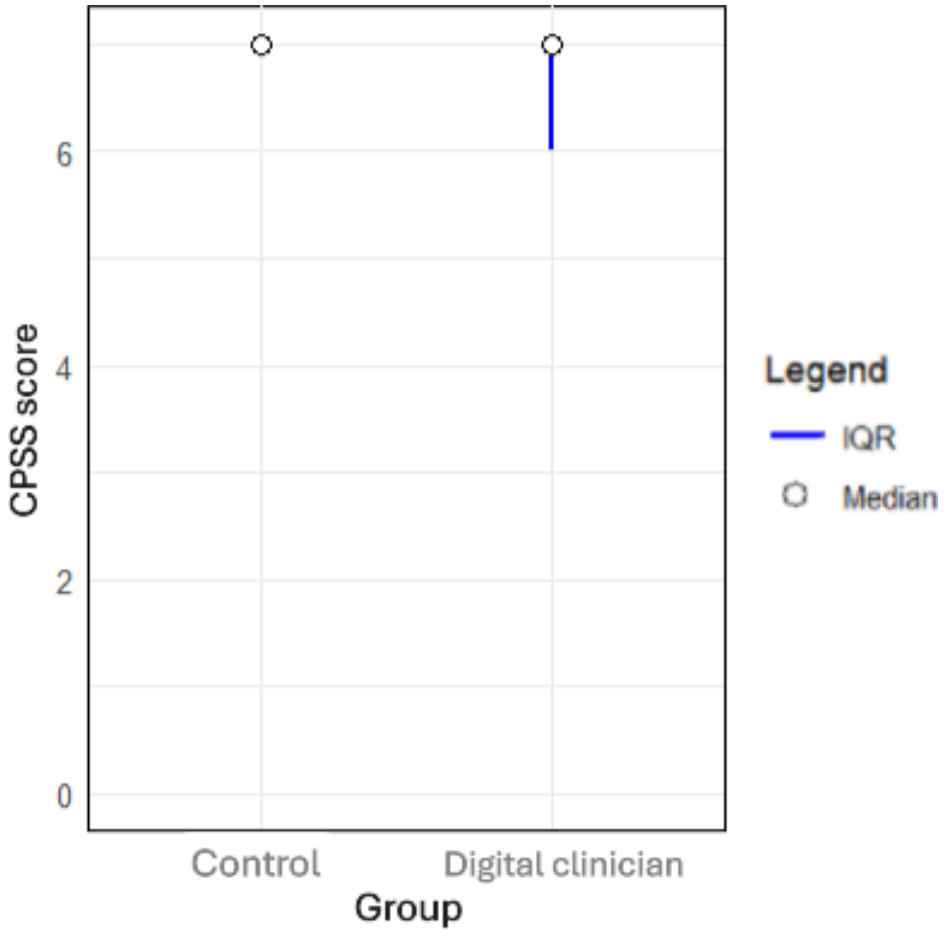
Table 3. Knowledge, trust, and satisfaction levels were assessed in participants that received their education from both the "digital clinician" and the clinical nurse specialist (controls).

Secondary Outcome	Digital clinician (n=27), median (IQR)	Control (n=16), median (IQR)	P value
Knowledge	10.0 (10.0-11.0)	8.0 (7.0-9.3)	<.001
Trust-Distrust Scale Domain			
I am confident of the nurse's knowledge and skills	7.0 (7.0-7.0)	7.0 (7.0-7.0)	.18
The nurse cares about you as a person	6.0 (4.0-7.0)	7.0 (7.0-7.0)	.003
I am (not) suspicious of the nurse's intentions and actions	5.0 (1.5-7.0)	7.0 (6.0-7.0)	.002
Overall Trust	7.0 (4.8-7.0)	7.0 (7.0-7.0)	<.001
Consultation Satisfaction			
Overall Satisfaction	7.0 (6.0-7.0)	7.0 (7.0-7.0)	<.001

Consultation Satisfaction

In total, 44% (12/27) respondents in the "digital clinician" group were more than 90% satisfied. The lowest score was 63%. In the control group, 94% (15/16) respondents were more than 90% satisfied with their consultation. The lowest score was 86%. The IQR of the "digital clinician group" was 6-7 (median 7), whereas it was 7-7 (median 7) in the control group (Mann-Whitney P<.001). This is shown in Figure 4 and Table 3 on a Likert scale between 1 and 7.

Figure 4. Satisfaction scores across control and digital clinician groups. CPSS: Overall Satisfaction with Primary Care Physicians Scale

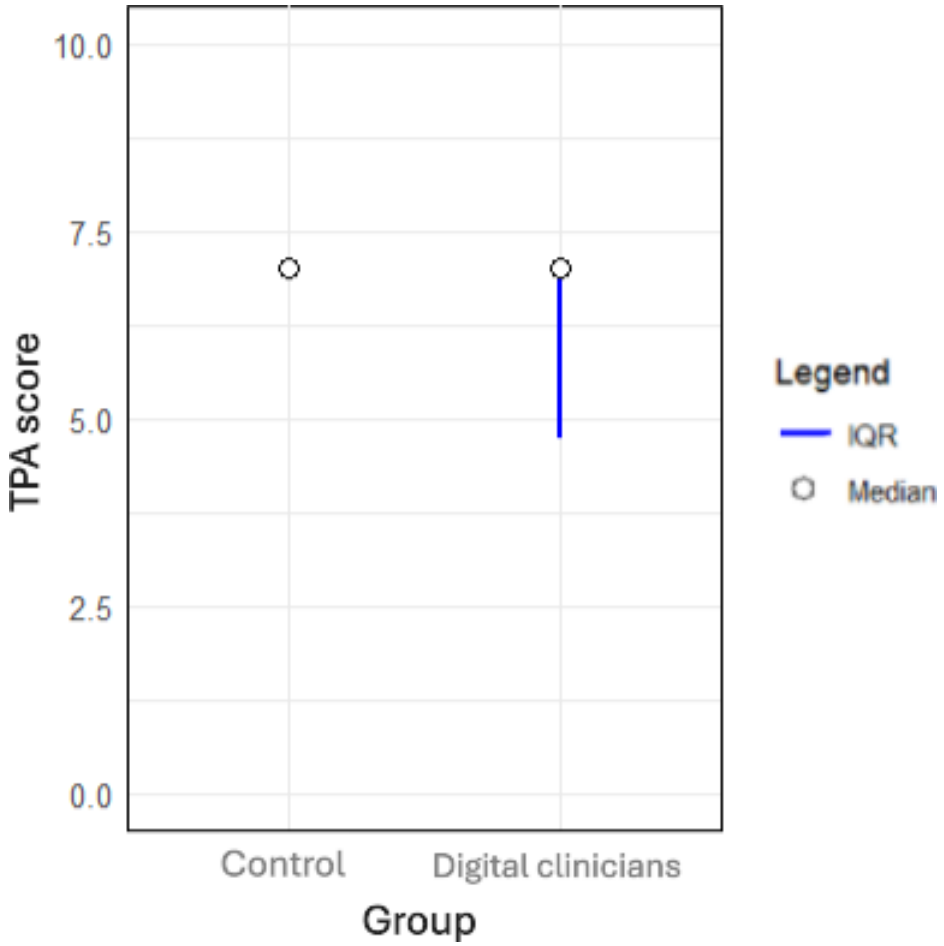


Trust

Figure 5 shows that participants in the "digital clinician" group scored their education provider lower for empathy (P=.003), clear intentions (P=.002), and overall trust (P<.001).

While Figures 4 and 5 show that the medians of the ordinal data are the same, the Mann Whitney U test is nonparametric and is not dependent on central tendency. As nonparametric data can be asymmetrical by nature, central tendency can be misleading. While the medians are the same, the IQR and ranks of the groups are shown to be significantly different with a P value <.05.

Figure 5. Trust scores across control and digital clinician groups. TPA: Trust between People and Automation Scale.



Resource Quality

On a scale of 100, the "digital clinician" was found to be 86.667 usable (Table 4), scoring lowest for interface quality and highest for interaction quality.

Of the 27 participants in the intervention group, 22 (81%) of participants affirmed they would use the "digital clinician" for educational purposes in their own time, while 2 (7%) of participants recorded that "maybe" they would. A total of 3 (11%) said they would not.

Table 4. The Usability of the "digital clinician" was rated by the user across three domains, scoring highest for interaction quality and ease of use, with 82% of users reporting they would use the resource in their own time.

Usability Domain	Values, mean (SD)
Interaction quality	89.6 (15.1)
Interface quality	81.5 (17.4)
Ease of use	88.9 (16.0)
Overall	86.7 (16.4)
Would you use this resource in your own time? n (%)	
Yes	22 (82)
No	3 (11)
Maybe	2 (7)

Analysis of Open-Ended Feedback

Below are the described questions about the analysis of open-ended feedback.

Question: "What do you think about using avatars with automated conversation in a healthcare environment?"

Four responses made reference to a great invention or great possibility for future use. Seven said the idea was good, or very good. Four described it as "okay." One respondent felt the avatar was "not ideal," and another stated they would "prefer a human." Two responses were of mixed sentiment:

"I don't think the real nurse can be replaced but it is great tool for people who live far away to refer to"… "They are helpful but. They are not as good as having a human explaining things to you."

Question: "What concerns would you have about using avatars with automated conversation like this in the future?"

Twelve responded that they would have no issues using the resource. Two expressed concerns over technical issues, such as freezing and audio quality. One expressed concern over language barriers. Four expressed concerns regarding the lack of empathy or possible loss of personal touch. One expressed concern over the ability of the avatar to answer the questions required. In total, 20%-25% expressed some form of apprehension. The prevalence of triggering the wrong conversational stream was not formally measured in the study; however, users did not remark on it when providing open-ended feedback on usability.

Question: "Any other comments?"

Two responses:

good to use at home

Needle testing could have been shown twice

Missing Data

Those with complete data had higher levels of education, female sex, were less trusting of (P=.04) and less satisfied (P=.03) with their health care professional than those lost to follow-up (Multimedia Appendix 3). However, the response rates of those in "digital clinician" and the "control" groups at follow-up were similar, as were measures of knowledge, self-efficacy, and usability.

Discussion

Principal Findings

Those who received education from the digital clinician were significantly more knowledgeable about their medication and its administration. They tended to have less stress and anxiety associated with using their injections, though this was not significant. Paradoxically, they had less confidence administering injections. They also trusted that their education provider was significantly less than those in the control group and were significantly less satisfied with their consultation. This suggests that from a health care psychology standpoint, participants' injection-related confidence was more related to who they received their information from, rather than how well informed they were.

While being less satisfied than those in the control group, the intervention group still reported very high levels of trust and satisfaction at levels, with median levels of 7/7 in both measures. Participants were extremely positive about the intervention, with more than 80% of participants expressing that they would use the resource in their own time. If not used clinically, distribution of a "digital clinician" for home use could also add value. The study did not recruit adequate numbers to test noninferiority at the predetermined level of self-efficacy due to global semaglutide shortages hindering patient access. However, there were significant differences in a range of secondary outcomes.

Human clinician-patient interactions are not scripted. They are shaped by human factors, including rapport and variability. Consultations with the "digital clinician" are more uniform and consistent. This may explain why a human was more successful at ensuring trust and satisfaction, while the "digital clinician" was more effective at ensuring information was provided, tested, and retained.

Certain studies have compared physicians versus chatbots, in educational settings, by assessing accuracy. No study has been uncovered showing that patients themselves were better educated by a chatbot or an LLM than by a specialist clinician. It is one thing to assess what information the chatbot is providing; it is another thing to assess if a patient is receiving, retaining, and trusting that information. Moving forward, only the true benefit of "digital clinicians" can be judged when research examines the risks and benefits of integration on a systemic level.

Standards for new technologies should be established to enable regulation, safety, and trust. Bespoke task-specific conversation streams such as this, which mimic large language models, may be easier to control and regulate while taking advantage of growing trust and recognition in the public domain.

Automation may incorporate tradeoffs in patient satisfaction and health care system trust, but may offer benefits in comprehensiveness, efficacy, and uniformity. Tradeoffs could potentially be minimized through research, human oversight, and clear physician accountability. The opportunity cost of using these technologies or not should be considered in terms of productivity, finance, access, and workforce and patient health.

Limitations

Limitations of the study include the loss of participants to follow-up questioning due to a global shortage of semaglutide, possibly contributing to some of the nonsignificant results in the study. There were significant missing data at the 2-week follow-up, which were omitted from analysis. Table S1 in Multimedia Appendix 4 shows the different profile of responders at the 2-week follow-up. Responders tended to have stronger views and were not a reliable representation of the initial cohort, making the conclusions regarding self-efficacy less applicable to the general population.

Nonvalidated measures of knowledge and adapted measures of usability, trust, and consultation satisfaction were used. While the trial design focuses on intergroup analyses, nonvalidated measures preclude direct comparison of the study cohort with the study population, as existing data for the latter does not exist using the same parameters. As a

result, nonvalidated and adapted measures may not correlate well with other health outcomes such as compliance or the incidence of side effects as established measures.

This feasibility trial, while not measuring traditional health outcomes, would have been strengthened by prospective trial registration, a prerequisite for all interventional AI trials from January 01, 2025 [38]. Trial registration was initially omitted from the design process. Contributing factors included the novelty of the intervention and the relative sparsity of an established industry proforma at the time. During the trial and data analysis period, further recognition of AI use as a significant health care intervention led to the emergence of industry directives, prompting reflection and retrospective trial registration. While not included in this study, further longitudinal research could focus on clinical outcomes such as adherence to therapy or weight loss over longer periods (eg, 1 year).

Conclusion

There is a worsening shortage of health care workers globally. Systems are not adapting to keep up with demand in areas such as bariatric medicine. "Digital clinicians,"chatbots integrated with digital avatar and emotional intelligence technology, may be part of scalable AI solutions.

"Digital clinicians" may ensure higher levels of information retention among patients compared with humans. However, users have significantly lower levels of trust and satisfaction in their information provider. Despite not being as satisfied, users are still overwhelmingly positive about their consultation, with over 80% saying they would use the resource in their own time.

Offering a degree of automation feasibly allows existing health care workers to focus on more demanding tasks, reduce stress, and improve health care system performance. It may offer more patients and systems access to medication and resources, provided they have the necessary internet connection and operability.

Safety should be established through regulation and research to improve trust and satisfaction. Ethical principles such as human oversight and accountability should build on this. Rigorous policy analysis should be performed to assess the potential tradeoffs in using these technologies as AI becomes more widespread across resource-rich and resource-poor contexts.

Conflicts of Interest

Kate Loveys is a former employee of and contractor to Soul Machines. Elizabeth Broadbent is a former contractor to Soul Machines. Francis Finucane is an investigator (unpaid) on the REDEFINE 2 and REDEFINE 3 randomized controlled trials, run by Novo Nordisk. He is a paid member of the Data Safety and Monitoring Board for the LEGEND and LEAP randomised controlled trials, run by the University of Michigan. He was a paid reviewer for the Danish Diabetes Academy until 2024.

Multimedia Appendix 1

Video of digital clinician in use.
[MP4 File (MP4 video File), 204027 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Outcome measures.
[DOCX File (Microsoft Word File), 66 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Trial protocol.
[PDF File (Adobe File), 137 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Self-efficacy scores across control and digital clinician groups.
[PDF File (Adobe File), 175 KB-Multimedia Appendix 4]

Checklist 1

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe File), 12519 KB-Checklist 1]

References

1. Obesity. World Health Organization. 2022. URL: https://www.who.int/health-topics/obesity#tab=tab_1 [Accessed 2025-06-12]
2. WHO European regional obesity report 2022. WHO Regional Office for Europe. 2022. URL: https://www.who.int/europe/publications/i/item/9789289057738 [Accessed 2025-07-17]
3. Flegal KM, Panagiotou OA, Graubard BI. Estimating population attributable fractions to quantify the health burden of obesity. Ann Epidemiol. Mar 2015;25(3):201-207. [doi: 10.1016/j.annepidem.2014.11.010] [Medline: 25511307]

4. Puhl R, Suh Y. Health consequences of weight stigma: implications for obesity prevention and treatment. Curr Obes Rep. Jun 2015;4(2):182-190. [doi: 10.1007/s13679-015-0153-z] [Medline: 26627213]
5. Jastreboff AM, Kaplan LM, Frías JP, et al. Triple-hormone-receptor agonist retatrutide for obesity - a phase 2 trial. N Engl J Med. Aug 10, 2023;389(6):514-526. [doi: 10.1056/NEJMoa2301972] [Medline: 37366315]
6. Stevens ER, Elmaleh-Sachs A, Lofton H, Mann DM. Lightening the load: generative AI to mitigate the burden of the new era of obesity medical therapy. JMIR Diabetes. Nov 14, 2024;9:e58680. [doi: 10.2196/58680] [Medline: 39622675]
7. Cecchini M. Use of healthcare services and expenditure in the US in 2025: The effect of obesity and morbid obesity. PLoS One. 2018;13(11):e0206703. [doi: 10.1371/journal.pone.0206703] [Medline: 30403716]
8. Arredondo K, Touchett HN, Khan S, Vincenti M, Watts BV. Current programs and incentives to overcome rural physician shortages in the United States: a narrative review. J Gen Intern Med. Jul 2023;38(Suppl 3):916-922. [doi: 10.1007/s11606-023-08122-6] [Medline: 37340266]
9. Zurynski Y, Herkes-Deane J, Holt J, et al. How can the healthcare system deliver sustainable performance? A scoping review. BMJ Open. May 24, 2022;12(5):e059207. [doi: 10.1136/bmjopen-2021-059207] [Medline: 35613812]
10. Workforce Health And Productivity. Health Aff (Millwood). Feb 2017;36(2):200-201. [doi: 10.1377/hlthaff.2016.1580]
11. Vrhovec J, Tajnikar M. Population ageing and healthcare demand: The case of Slovenia. Health Policy. Nov 2016;120(11):1329-1336. [doi: 10.1016/j.healthpol.2016.09.007] [Medline: 28340936]
12. Doleman G, De Leo A, Bloxsome D. The impact of pandemics on healthcare providers' workloads: A scoping review. J Adv Nurs. Dec 2023;79(12):4434-4454. [doi: 10.1111/jan.15690] [Medline: 37203285]
13. Sarkar S. "Past the point of sanity"-South East Asia faces critical shortage of healthcare workers. BMJ. Aug 9, 2023;382:1655. [doi: 10.1136/bmj.p1655] [Medline: 37558238]
14. Leong SL, Teoh SL, Fun WH, Lee SWH. Task shifting in primary care to tackle healthcare worker shortages: An umbrella review. Eur J Gen Pract. Dec 2021;27(1):198-210. [doi: 10.1080/13814788.2021.1954616] [Medline: 34334095]
15. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: a new era of artificial intelligence in medicine. Ir J Med Sci. Dec 2023;192(6):3197-3200. [doi: 10.1007/s11845-023-03377-8] [Medline: 37076707]
16. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). Mar 19, 2023;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
17. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. Jun 1, 2023;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
18. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health. Feb 15, 2023;20(4):3378. [doi: 10.3390/ijerph20043378] [Medline: 36834073]
19. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. NEJM AI. Jan 2024;1(1). [doi: 10.1056/AIp2300031]
20. Harrison RM, Lapteva E, Bibin A. Behavioral nudging with generative AI for content development in SMS health care interventions: case study. JMIR AI. Oct 15, 2024;3:e52974. [doi: 10.2196/52974] [Medline: 39405108]
21. Dorr DA, Adams L, Embí P. Harnessing the promise of artificial intelligence responsibly. JAMA. Apr 25, 2023;329(16):1347-1348. [doi: 10.1001/jama.2023.2771] [Medline: 36972068]
22. Chandra S, Ward P, Mohammadnezhad M. Investigating patient trust in doctors: a cross-sectional survey of out-patient departments in Fiji. Int Q Community Health Educ. Jul 2021;41(4):369-377. [doi: 10.1177/0272684X20967602] [Medline: 33086937]
23. Zahed K, Mehta R, Erraguntla M, Qaraqe K, Sasangohar F. Understanding patient beliefs in using technology to manage diabetes: path analysis model from a national web-based sample. JMIR Diabetes. May 3, 2023;8:e41501. [doi: 10.2196/41501] [Medline: 37133906]
24. Curtis RG, Bartel B, Ferguson T, et al. Improving user experience of virtual health assistants: scoping review. J Med Internet Res. Dec 21, 2021;23(12):e31737. [doi: 10.2196/31737] [Medline: 34931997]
25. Zemore SE. The effect of social desirability on reported motivation, substance use severity, and treatment attendance. J Subst Abuse Treat. Jun 2012;42(4):400-412. [doi: 10.1016/j.jsat.2011.09.013] [Medline: 22119180]
26. Adong J, Fatch R, Emenyonu NI, et al. Social desirability bias impacts self-reported alcohol use among persons with HIV in Uganda. Alcohol Clin Exp Res. Dec 2019;43(12):2591-2598. [doi: 10.1111/acer.14218] [Medline: 31610017]

27. Chattopadhyay D, Ma T, Sharifi H, Martyn-Nemeth P. Computer-controlled virtual humans in patient-facing systems: systematic review and meta-analysis. J Med Internet Res. Jul 30, 2020;22(7):e18839. [doi: 10.2196/18839] [Medline: 32729837]
28. Altman DG, Bland JM. Treatment allocation by minimisation. BMJ. Apr 9, 2005;330(7495):843. [doi: 10.1136/bmj.330.7495.843]
29. Keininger D, Coteur G. Assessment of self-injection experience in patients with rheumatoid arthritis: psychometric validation of the Self-Injection Assessment Questionnaire (SIAQ). Health Qual Life Outcomes. Jan 13, 2011;9(2):MC3027089. [doi: 10.1186/1477-7525-9-2] [Medline: 21232106]
30. Pompilus F, Ciesluk A, Strzok S, et al. Development and psychometric evaluation of the assessment of self-injection questionnaire: an adaptation of the self-injection assessment questionnaire. Health Qual Life Outcomes. Nov 4, 2020;18(1):355. [doi: 10.1186/s12955-020-01606-7] [Medline: 33148261]
31. Hojat M, Louis DZ, Maxwell K, Markham FW, Wender RC, Gonnella JS. A brief instrument to measure patients' overall satisfaction with primary care physicians. Fam Med. Jun 2011;43(6):412-417. [Medline: 21656396]
32. Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon. Mar 2000;4(1):53-71. [doi: 10.1207/S15327566IJCE0401_04]
33. Hajesmaeel-Gohari S, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. BMC Med Inform Decis Mak. Feb 2, 2021;21(1):36. [doi: 10.1186/s12911-021-01407-y] [Medline: 33531013]
34. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. Sep 2020;26(9):1364-1374. [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
35. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ. Sep 9, 2020;370:m3210. [doi: 10.1136/bmj.m3210] [Medline: 32907797]
36. The jamovi project. jamovi (Version 24). 2023. URL: https://www.jamovi.org [Accessed 2025-07-09]
37. R core team. R: A Language and environment for statistical computing (Version 41) [Computer software]. 2022. URL: https://cran.r-project.org [Accessed 2025-07-17]
38. Drazen JM, Haug CJ. Trials of AI interventions must be preregistered. NEJM AI. Mar 28, 2024;1(4):AIe2400146. [doi: 10.1056/AIe2400146]

Abbreviations

AI: artificial intelligence
CONSORT-AI: Consolidated Standards of Reporting Trials- Artificial Intelligence
SIAQ: Self-Injection Assessment Questionnaire
SPIRIT-AI: S
SPIRIT-AI : Standard Protocol Items: Recommendations for Interventional Trials involving Artificial Intelligence
TUQ: Technology Usability Questionnaire
WHO: World Health Organization

Edited by Sheyu Li; peer-reviewed by Hengwei Zhang, Simon Laplante; submitted 21.06.2024; final revised version received 19.03.2025; accepted 25.04.2025; published 30.07.2025

Please cite as:
Coleman S, Lynch C, Worlikar H, Kelly E, Loveys K, Simpkin AJ, Walsh JC, Broadbent E, Finucane FM, O' Keeffe D "Digital Clinicians" Performing Obesity Medication Self-Injection Education: Feasibility Randomized Controlled Trial JMIR Diabetes 2025;10:e63503
URL: https://diabetes.jmir.org/2025/1/e63503
doi: 10.2196/63503