

“Digital Clinicians” performing OBESITY Medication Self-Injection Education: A Feasibility Randomised Controlled Trial

Sean Coleman, Caitríona Lynch, Hemendra Worlikar, Emily Kelly, Kate Loveys, Andrew J Simpkin, Jane C Walsh, Elizabeth Broadbent, Francis M Finucane, Derek O' Keeffe

Submitted to: JMIR Diabetes
on: June 21, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	22
Figures	23
Figure 1.....	24
Figure 2.....	25
Figure 3.....	26
Figure 4.....	27
Multimedia Appendixes	28
Multimedia Appendix 0.....	29
Multimedia Appendix 0.....	29
Multimedia Appendix 0.....	29
Multimedia Appendix 1.....	29
Multimedia Appendix 2.....	29
CONSORT (or other) checklists.....	30
CONSORT (or other) checklist 0.....	30
CONSORT (or other) checklist 0.....	30

“Digital Clinicians” performing OBESITY Medication Self-Injection Education: A Feasibility Randomised Controlled Trial

Sean Coleman^{1,2} MB, Bch, BAO, MSc; Caitríona Lynch² BSc, MSc; Hemendra Worlikar¹ BEng, MEng; Emily Kelly¹; Kate Loveys³ PhD; Andrew J Simpkin¹ PhD; Jane C Walsh¹ PhD; Elizabeth Broadbent³ PhD; Francis M Finucane^{1,2} MD; Derek O' Keeffe^{1,2} PhD, MD, MBA, MEng

¹ University of Galway Galway IE

² University Hospital Galway Galway IE

³ University of Auckland Auckland NZ

Corresponding Author:

Sean Coleman MB, Bch, BAO, MSc

University Hospital Galway

University Hospital Galway

Newcastle

Galway

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 investigates the use of an AI 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 centre.

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

Results: 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 post-consultation ($p<0.001$) and had better feelings about their injections ($p=.079$). Patients in the control group were more satisfied with their consultation ($p<0.001$) and had more trust in their education provider ($p<0.001$). There was no significant difference in reported levels of self-efficacy. 81% of participants 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 healthcare 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 healthcare systems. The extent to which remains to be determined in future studies. Clinical Trial: ISRCTN Registry ISRCTN12345678; <https://www.isrctn.com/ISRCTN12382879>

(JMIR Preprints 21/06/2024:63503)

DOI: <https://doi.org/10.2196/preprints.63503>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org](#)

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in [https://www.jmir.org](#)

Original Manuscript

“Digital Clinicians” performing OBESITY Medication Self-Injection Education: A Feasibility Randomised Controlled Trial

S Coleman^{1,2}, C Lynch², H Worlikar¹, E Kelly¹, K Loveys³, AJ Simpkin¹, JC Walsh¹, E Broadbent³,
FM Finucane^{1,2}, D O’ Keeffe^{1,2}

¹University of Galway, Ireland.

²University Hospital Galway.

³University of Auckland, New Zealand.

Word Count: 3838

References: 35

Tables: 2

Figures: 4

Corresponding Author: Seán Coleman

Address: 12 Tegan Court, Screggan, Tullamore, Co. Offaly.

Telephone Number: +353851604880

Fax Number: + 0135391544544

E-mail Address for correspondence: 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.

Objectives: This study investigates 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 centre.

Methods: A “Digital Clinician” with facial and vocal recognition technology was generated with a bespoke 10-to-15-minute clinician-validated tutorial. A feasibility randomised controlled non-inferiority trial compared knowledge test scores, self-efficacy, consultation satisfaction, and trust levels between those using the AI-powered clinician avatar on-site and those receiving conventional semaglutide education from nursing staff. Attitudes were recorded immediately after the intervention and again at two-weeks after the education session.

Results: 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 post-consultation (IQR 10.0-11.0 vs 7.0-9.3, $p<0.001$). Patients in the control group were more satisfied with their consultation (IQR 6.0-7.0 vs 7.0-7.0, $p<0.001$) and had more trust in their education provider (IQR 4.8-7.0 vs 7.0-7.0, $p<0.001$). There was no significant difference in reported levels of self-efficacy ($p=.57$). 81% (22 of 27) of participants in the intervention group said they would use the resource in their own time.

Conclusion: Bespoke AI chatbots integrated with digital avatars to create a “digital clinician” may perform healthcare 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 healthcare systems. The extent to which remains to be determined in future studies.

Keywords: “Automation”, “Machine-Human Interface”, “Clinical Education”, “Digital clinician”, “Virtual Human”, “Chatbot”, “Trust”

Registration: ISRCTN ISRCTN12382879, <https://www.isrctn.com/ISRCTN12382879>

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 (GLP-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 WHO European Region [2] potentially eligible, utility is hindered by scalability of clinical services and supply. Generative AI has been identified to 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 pre-requisite at many clinics for the prescription of pharmacotherapy. Even in resource rich countries, the necessary services are not always available. Studies have predicted the U.S.A need careful restructuring of health service expenditure to meet demand for the costs of overweight and obesity services [7].

Healthcare systems are underresourced and understaffed, particularly in rural areas [8]. This impacts productivity, sustainability [9] and the health of both patient and healthcare professional [10]. Healthcare worker shortages, propelled by aging populations [11] and the Covid-19 pandemic [12] are exacerbated in areas such as the Global South, increasing healthcare inequities [13]. Tasks that healthcare workers perform are often repetitive and administrative. Redistribution of such tasks has been shown to potentially improve health outcomes such as blood pressure, HbA1c, 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 personalised medicine [16]. Anonymised AI chatbots have been judged to provide better, more concise, empathetic answers to general health queries than verified physicians [17]. Chat Generative Pre-Trained Transformer(GPT)-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 personalised messaging systems are being researched to enhance behavioural 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 behaviours, compliance and quality of life [22]. Willingness to use supportive technologies have 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, e.g. 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 standardisation, efficiency, effectiveness, cost, confidentiality and access. Social desirability response bias is associated with higher levels of treatment non-adherence [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 healthcare professional.

Virtual humans have been shown to be efficacious in non-clinical 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

A randomised controlled non-inferiority trial of “digital clinician”-led patient education was designed with ethical approval from Galway University Hospitals (GUH), Clinical Research Ethics Committee (Ref: CA 2920). The trial and trial protocol were guided by SPIRIT-AI[28] and CONSORT-AI[29] extensions.

Further information on the study protocol can be found in the supplementary appendix.

All patients at GUH 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/m².

Eligibility criteria included being over 18 years of age, 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, behaviour 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 were included. Interactive questions were used to maximise 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 so, 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 were used to design the conversational flow. Rigorous internal testing including over 100 iterations of the conversational flow were 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 device.

Two multi-scene videos 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 accessed online 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 behaviour. 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 healthcare institutions were included on the User Interface Screen to increase trust. Cinematic Screening offered different angles of the avatars face to make for a more interesting experience.

Multiple avatars of different genders, race and name 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.

All apps used were from third parties, with no affiliations to the research team.

Randomisation: Minimisation

Minimisation is a method employed in clinical trials and experimental research to allocate participants.[30][4] . Participants were allocated in order of enrolment, to the study arm which minimised the difference across four selected numerical baseline variables- Age, BMI, pre-tutorial knowledge of semaglutide score and pre-tutorial injection self-efficacy score. Order of enrolment in the study was determined by check-in time at the clinic. Allocation by minimisation 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 online suppository. The assessor then communicated the allocation to the research team on site.

Self-Efficacy

Self-efficacy was measured using the two-part validated Self Injection Assessment Questionnaire (SIAQ) [31]. The pre-injection SIAQ was used as baseline data.

The post-injection SIAQ was scored across five domains two-weeks post-intervention- 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's alpha and the test-retest coefficient were > 0.70 for all domains. [32].

Knowledge Attainment

Unvalidated knowledge assessments were designed specifically for the purpose of this research.

The pre-educational assessment tool consisted of four questions. The post-educational assessment

tool contained the same initial four questions, and a further eight 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 [33] has Criterion-related validity coefficients mostly in the 0.80s and 0.90s when considering empathy, physician recommendation and general satisfaction. Cronbach's coefficient alpha 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" has undergone validation [34] and been in use for over 20 years. It was adapted and integrated with the consultation satisfaction measure for this study.

Technology Usability and Future Use (TUQ)

The validated TUQ [35] was adapted. The questionnaire has high reliability with Cronbach Coefficient Alpha's over 0.8 across all domains, and high validity expressed through its multiple native questionnaires (35). A shortened version including questions from four of five 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 towards the "digital clinician".

Sample Size Calculations

When calculating sample size, a similar study using the validated Self-Injection Assessment Questionnaire (SIAQ) to measure self-efficacy was used (29). It showed a Standard Deviation of 1.5, on a 10 point-likert scale. A significant difference of two-thirds the standard deviation was proposed to boundary non-inferiority, with 80% power and alpha equals to 0.05, this would require 28 participants to be recruited into each arm of the study.

Ethical considerations

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 had received safety before being discharged with a medication information leaflet. Contact with the nurse minimised disruption of the patient-clinician relationship. This established human oversight and accountability.

Statistical Analysis and Missing Data

Data was analysed using Jamovi version 2.4.14 and R Studio [36, 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-to-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 two groups. P-values below 0.05 were deemed statistically significant. Conventional Content analysis was used to interpret qualitative data. Missing outcome data were omitted from analysis. Observed outcome data were analysed as if representative of the entire cohort.

Knowledge, Satisfaction, Trust and Usability were measured immediately post-tutorial. Subsequent

loss to follow-up at two-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 was explored in Table 2.



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. Of this cohort, 43 agreed to participate in the study. Of those enrolled, 100% (43 of 43) completed the pre-tutorial questionnaire, their assigned education session and their post-tutorial questionnaire. Only 18 participants, 41.9%, completed the two-week follow-up self-efficacy questionnaire.

See *Table 1*.

Self Efficacy

Both groups had a median self-efficacy level of 10.0/10.0, with the “digital clinician” group having a marginally lower first quartile of 8.0, compared to the control group (8.2)-see *Figure 1*. This difference was not statistically significant ($p=.517$).

Responses recording “Pain and Skin reactions” from the control group had a favourable, non-significantly higher mean rank.

Knowledge (Test Score)

Overall, the median test score of participants post-consultation in the “digital clinician” group was 10.0 out of 12 (84%), while median score in the control group was 8.0 out of 12 (69%) ($P<.001$). See *Figure 2*.

Consultation Satisfaction

44% (12 of 27) of respondents in the “digital clinician” group were over 90% satisfied. The lowest score was 63%.

In the Control group, 94% of (15 of 16) respondents were over 90% satisfied with their consultation. The lowest score was 86%.

The interquartile range of the “digital clinician group” was 6.0 to 7.0, whereas it was 7.0 to 7.0 in the control group (Mann-Whitney $p < .001$).

This is shown in *Figure 3* and *Table 1* on a likert scale between 1 and 7.

Trust

Figure 4 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 3 and 4* show that the medians of the ordinal data are the same, the Mann Whitney U test is non-parametric and is not dependent on central tendency. As non-parametric data can be asymmetrical by nature, central tendency can be misleading.

While the medians are the same, the interquartile ranges and ranks of the groups are shown to be significantly different with a p value $<.05$.

Resource Quality

On a scale of 100, the “digital clinician” was found to be 86.667 usable, scoring lowest for interface quality and highest for interaction quality.

3.2.5.2 Home Use
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. 3 (11%)] said they would not.

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. 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

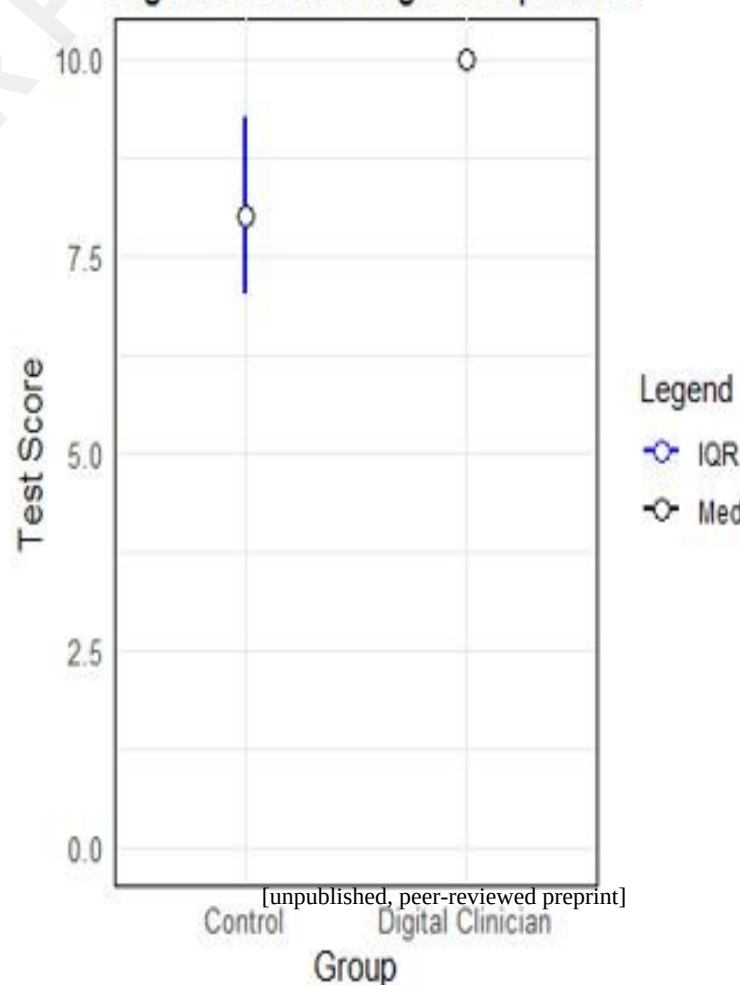
Those with complete data had higher levels of education, female gender, were less trusting of

Figure 1: Self-Efficacy Comparison



Mann-Whitney U test used to calculate p-value

Figure 2: Knowledge Comparison



[unpublished, peer-reviewed preprint]

Figure 3: Satisfaction Comparison

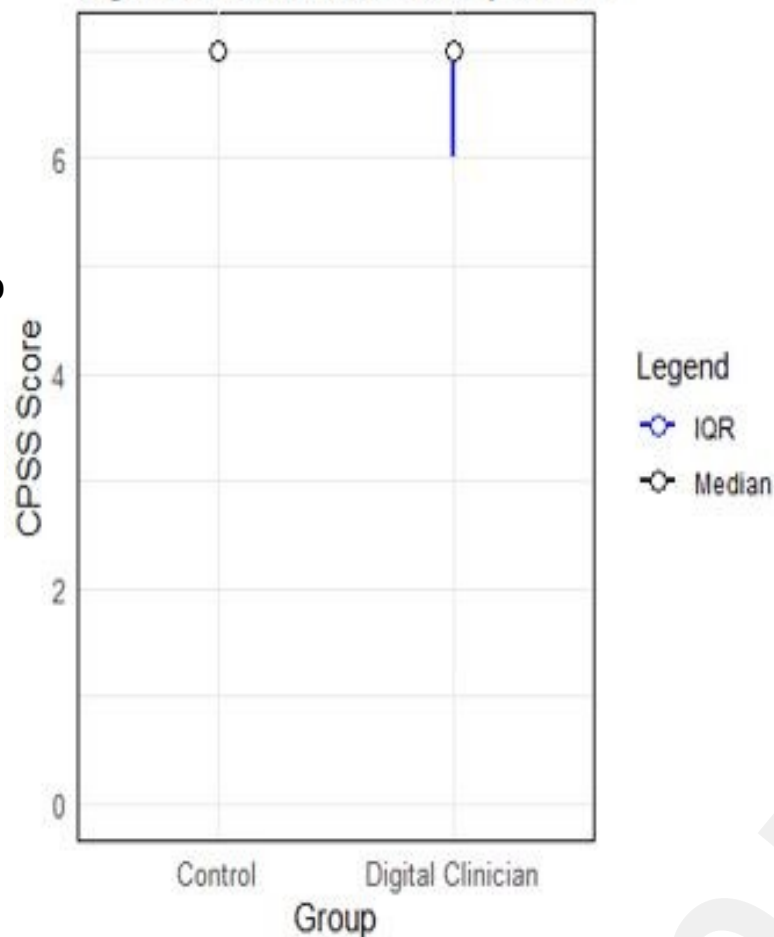
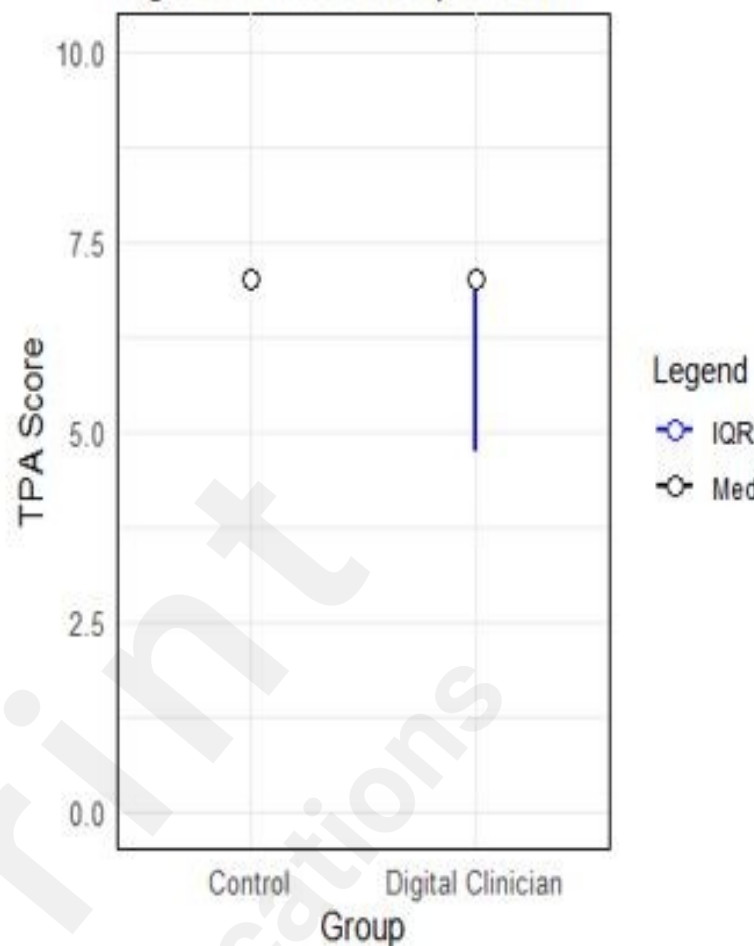


Figure 4: Trust Comparison



person

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

Usability

Interaction Quality

89.6

Interface Quality

81.5

Ease of Use

88.9

Overall

86.7

Would you use/ this Resource in your own time?

n(%)

Yes

22(81.5)

No

3(11.1)

Maybe

2(7.4)

Discussion

General

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 their education provider significantly less than those in the control group, and were significantly less satisfied with their consultation. This suggests that from a healthcare 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 7.0/7.0 in both measures. . Participants were extremely positive about the intervention with over 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 non-inferiority at the pre-determined 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're 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 that showed 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 trade-offs in patient satisfaction and healthcare system trust, but may offer benefits in comprehensiveness, efficacy and uniformity. Trade-offs could potentially be minimised 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 non-significant results in the study.. There were significant missing data at 2-weeks follow-up which were omitted from analysis. Table 2 shows the different profile of responders at two-weeks 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.

Non-validated measures of knowledge and adapted measures of usability, trust and consultation

satisfaction were used. While the trial design focuses on inter-group analyses, non-validated 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, non-validated and adapted measures may not correlate as 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 pre-requisite for all interventional AI trials from the 1st of January 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 healthcare 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 (e.g one year).

Conclusion

There is a worsening shortage of healthcare 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 to 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 healthcare workers to focus on more demanding tasks, reduce stress and improve healthcare 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 trade-offs in using these technologies as AI becomes more widespread across resource rich and resource poor contexts.

A short video of the “digital clinician” in use is included as a multimedia appendix as are the outcome measure tools.

References

1. Obesity [Internet]. World Health Organization. World Health Organization; 2022 [cited 2022Dec20]. Available from: https://www.who.int/health-topics/obesity#tab=tab_1.
2. WHO European Regional Obesity Report 2022. Copenhagen: WHO Regional Office for Europe; 2022. Licence: CC BY-NC-SA 3.0 IGO.
3. Flegal KM, Panagiotou OA, Graubard BI. Estimating population attributable fractions to quantify the health burden of obesity. *Ann Epidemiol.* Mar 2015;25(3):201-7. doi:10.1016/j.annepidem.2014.11.010 PMC4383168
4. Puhl R, Suh Y. Health Consequences of Weight Stigma: Implications for Obesity Prevention

and Treatment. *Curr Obes Rep.* Jun 2015;4(2):182-90. doi:10.1007/s13679-015-0153-z

5. Jastreboff AM, Kaplan LM, Frias JP, et al. Triple-Hormone-Receptor Agonist Retatrutide for Obesity - A Phase 2 Trial. *N Engl J Med.* Jun 26 2023;doi:10.1056/NEJMoa2301972

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 PMC11611792

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 PMC6221341

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 PMC10356718

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 PMC9125771

10. Workforce Health And Productivity. *Health Aff (Millwood).* Feb 1 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

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

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

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 PMC8330741

15. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci.* Apr 19 2023;doi:10.1007/s11845-023-03377-8

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)doi:10.3390/healthcare11060887 PMC10048148

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 PMC10148230

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)doi:10.3390/ijerph20043378 PMC9967747

19. Eriksen Alexander V, Möller S, Ryg J. Use of GPT-4 to Diagnose Complex Clinical Cases. *NEJM AI.* 2023/12/11 2023;1(1):AIp2300031. 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 PMC11522651

21. Dorr DA, Adams L, Embi P. Harnessing the Promise of Artificial Intelligence Responsibly. *JAMA.* Apr 25 2023;329(16):1347-1348. doi:10.1001/jama.2023.2771

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

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.

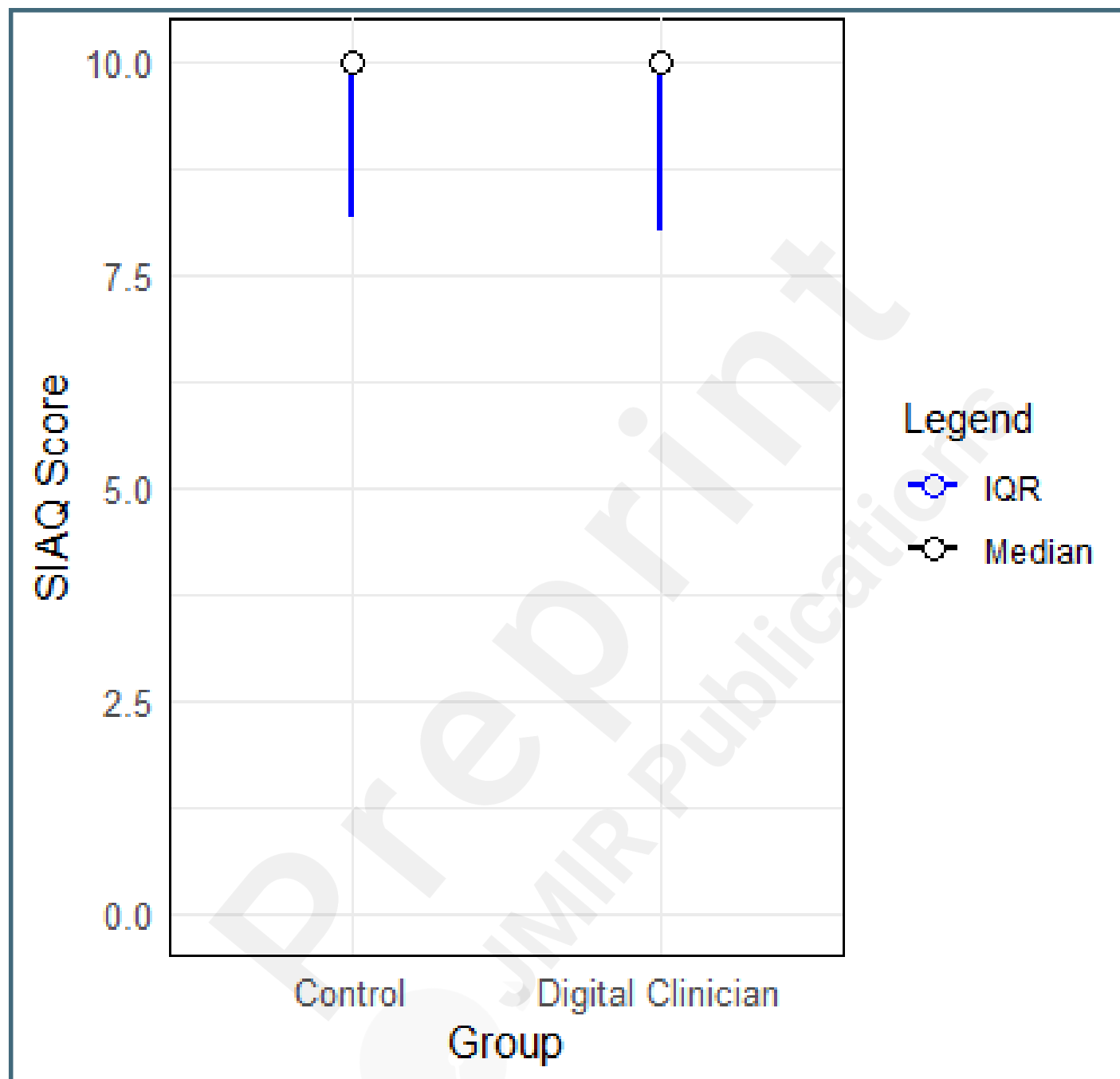
JMIR Diabetes. May 3 2023;8:e41501. doi:10.2196/41501 PMC10193211

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 PMC8734926
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):400-12. doi:10.1016/j.jsat.2011.09.013 PMC3296824
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 PMC7411366
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 PMC7426801
28. Rivera SC, Liu X, Chan AW, et al. 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 PMC7490785
29. Liu X, Cruz Rivera S, Moher D, et al. 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 PMC7598943
30. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. Apr 9 2005;330(7495):843. doi:10.1136/bmj.330.7495.843 PMC556084
31. 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. doi:10.1186/1477-7525-9-2 PMC3027089
32. 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 PMC7640481
33. 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):412-7.
34. Jiun-Yin Jian, Ann M. Bisantz & Colin G. Drury (2000) Foundations for an Empirically Determined Scale of Trust in Automated Systems, *International Journal of Cognitive Ergonomics*, 4:1, 53-71, DOI: 10.1207/S15327566IJCE0401_04.
35. 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 PMC7852181
36. The jamovi project (2023). jamovi. (Version 2.4) [Computer Software]. Retrieved from <https://www.jamovi.org>.
37. R Core Team (2022). R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2023-04-07).
38. Drazen Jeffrey M, Haug Charlotte J. Trials of AI Interventions Must Be Preregistered. *NEJM AI*. 0(0):AIe2400146. doi:10.1056/AIe2400146

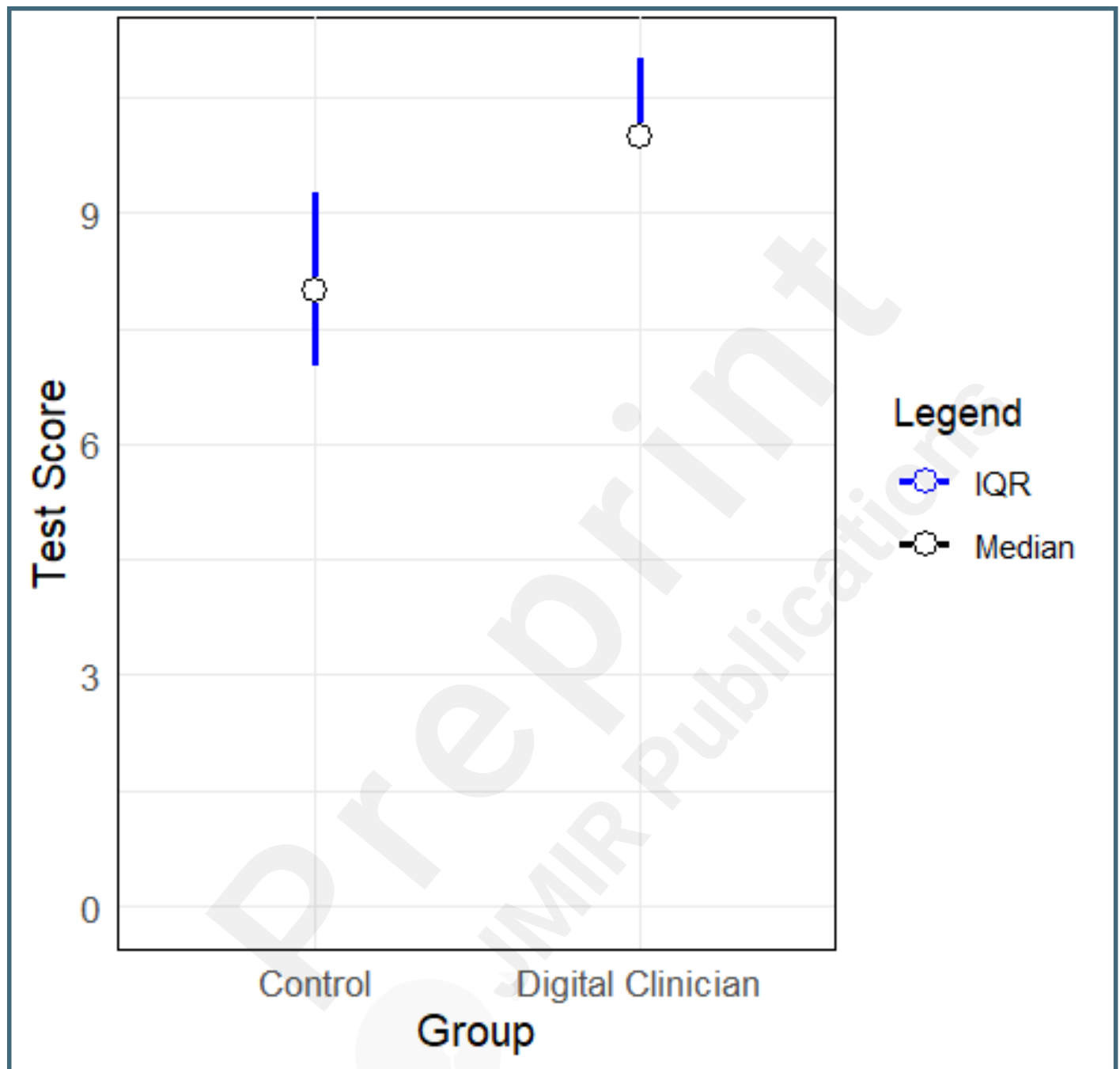
Supplementary Files

Figures

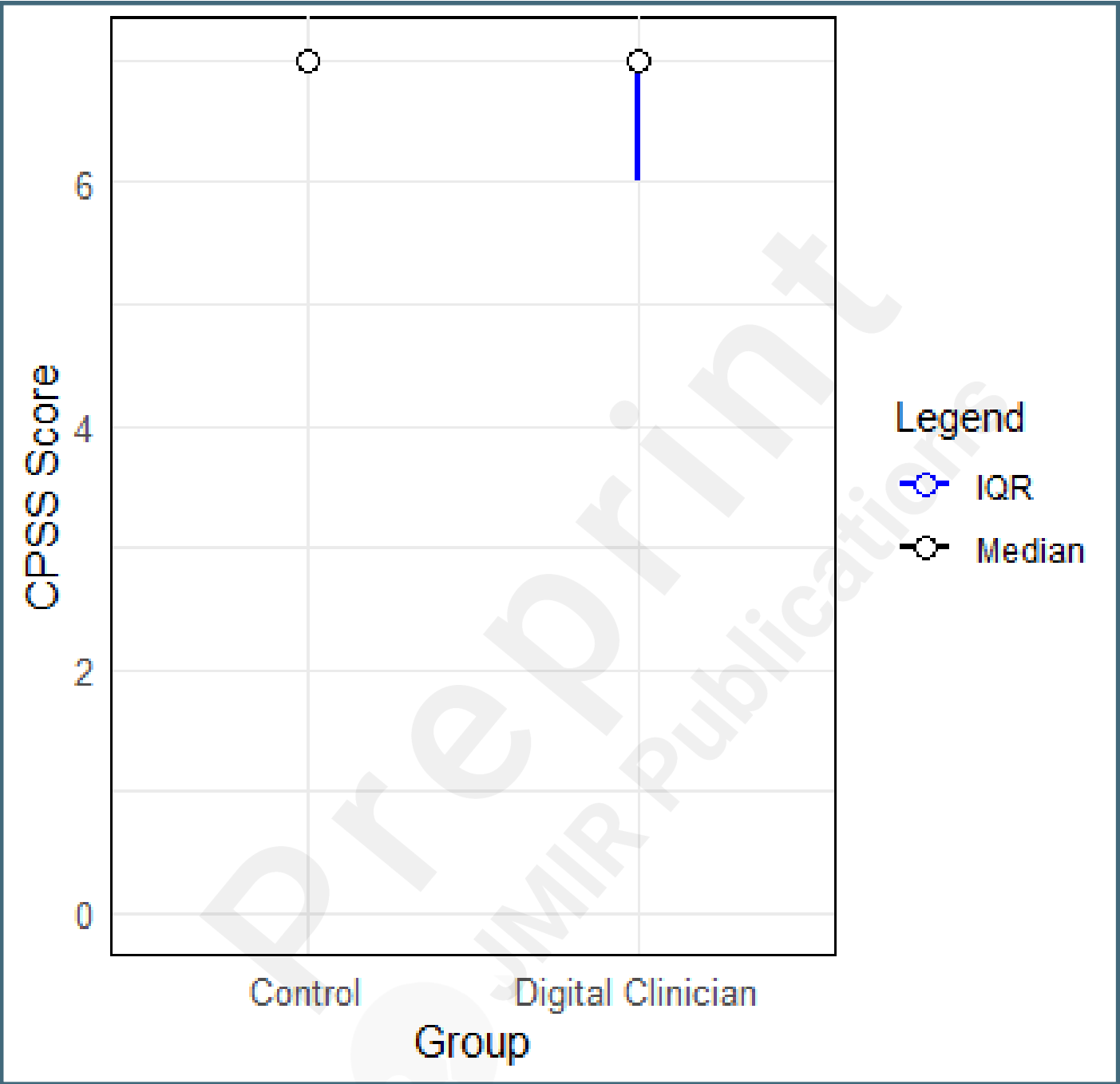
Self-Efficacy Scores across Control and Digital Clinician groups.



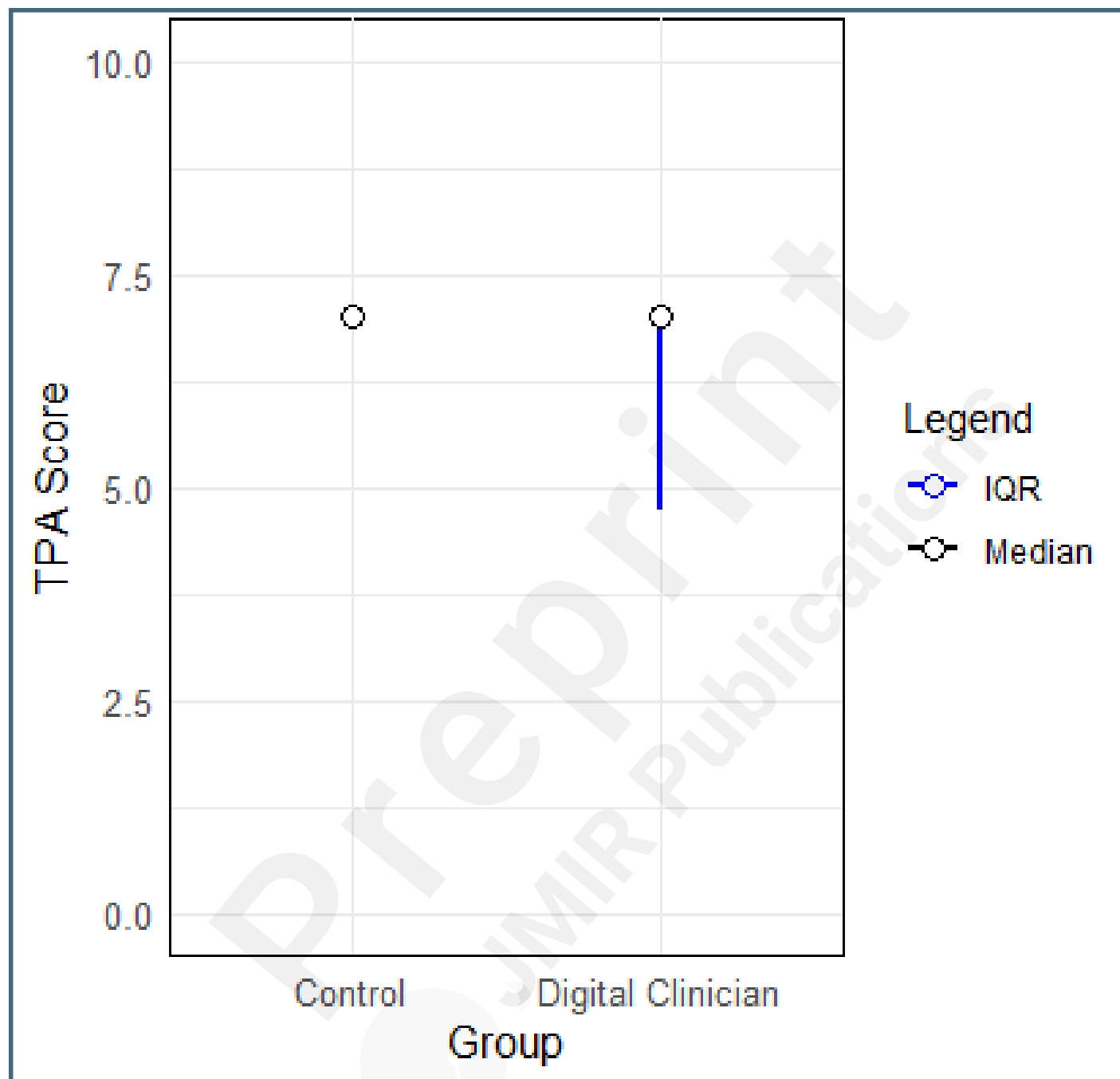
Knowledge Scores Across Control and Digital Clinician groups.



Satisfaction Scores Across Control and Digital Clinician groups.



Trust Scores Across Control and Digital Clinician groups.



Multimedia Appendixes

Trial Protocol.

URL: <http://asset.jmir.pub/assets/b172ad711ba787154c8af0a19d5cb522.pdf>

Video of "digital clinician".

URL: <http://asset.jmir.pub/assets/2d9b2635fadb2172a39cf6e1db884140.mp4>

Supplementary Appendix 3- Table2: Missing Data.

URL: <http://asset.jmir.pub/assets/30a7be289547047dc1879d947376c509.pdf>

Video of Digital Clinician in use.

URL: <http://asset.jmir.pub/assets/5c5e97c207f809160ba35685d52cb013.mp4>

Outcome Measures.

URL: <http://asset.jmir.pub/assets/2d298be1cd8ab83e0182942df7889029.docx>

CONSORT (or other) checklists

CONSORT EHealth Checklist.

URL: <http://asset.jmir.pub/assets/3cd4c7e6dadf01bae164cfe45fd91b62.pdf>

CONSORT EHealth checklist.

URL: <http://asset.jmir.pub/assets/59c0779d5452544ee290a381abc2db04.pdf>