

Impact of large language model Dialogue System on Perioperative Anxiety and Depression in Patients Undergoing Total Knee Arthroplasty: A Randomized Controlled Trial

Yixiang Zhang, Chunlei Wu, Tao Yang, Haotian Yin, Jingchao Du, Xin Fu, JUN LIU

Submitted to: JMIR Perioperative Medicine
on: April 06, 2025

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..... 25

 Figures 26

 Figure 1..... 27

 Figure 2..... 28

 Figure 3..... 29

Impact of large language model Dialogue System on Perioperative Anxiety and Depression in Patients Undergoing Total Knee Arthroplasty: A Randomized Controlled Trial

Yixiang Zhang^{1, 1*}; Chunlei Wu^{2*}; Tao Yang¹; Haotian Yin^{1, 3}; Jingchao Du¹; Xin Fu¹; JUN LIU¹

¹ Tianjin Hospital tianjin CN

² The united logistics security force of 983 hospital, Tianjin, China tianjin CN

³ Tianjin University tianjin CN

*these authors contributed equally

Corresponding Author:

JUN LIU

Tianjin Hospital

No. 406, Jiefang Nan Road, Tianjin, People's Republic of China, 300211

tianjin

CN

Abstract

Background: Total knee arthroplasty (TKA) is a widely performed procedure for knee osteoarthritis, yet patient satisfaction remains suboptimal, with preoperative anxiety and depression contributing to poor postoperative outcomes. Large-scale language models (LLMs), such as ChatGPT, have shown potential in healthcare applications but remain underexplored in the context of TKA. This study evaluates the impact of an AI-driven dialogue system on perioperative anxiety and depression in TKA patients.

Objective: This study aimed to evaluate the effect of LLMs dialogue system on improving anxiety and depression, as well as overall psychological well-being, in patients undergoing TKA. The primary objective was to assess the system's impact on reducing perioperative anxiety and depression levels before and after surgery. Secondary objectives included evaluating patient satisfaction, the feasibility and scalability of the AI system compared to traditional methods, and its potential for broader clinical application in surgical settings.

Methods: We expect to conduct a parallel randomized and controlled trial in Tianjin Hospital. This trial content needs to gather 210 participants between the ages of 50-80 and randomly divide them into two groups, namely, the LLM intervention group or the usual care group. ChatGPT 4.0 software was used to conduct a personalized psychological education activity for about 20-30 minutes for the LLM intervention group every day, and the routine nursing group was used as the control group, receiving standard preoperative education. The test results were evaluated by the hospital anxiety and Depression Scale (HADS) at baseline, 48 hours, 24 hours before surgery, and 24 hours, 48 hours, 72 hours after surgery to determine whether artificial intelligence is effective in anti anxiety and depression, and improve patients' surgical mood and health.

Results: Statistical analyses showed that compared with the usual care group, the reduction of anxiety and depression in the LLM group was significantly greater. We found that both groups of patients experienced significant anxiety and depression before surgery. With the intervention and guidance of artificial intelligence, these symptoms in the LLM group continued to alleviate after surgery. The effect was particularly significant at 48 and 72 hours after surgery ($P < .001$), while the conventional nursing group showed a slower decline.

Conclusions: Experimental data shows that an intelligent psychological support system driven by artificial intelligence can effectively reduce anxiety and depression in TKA patients before and after surgery. Compared with traditional psychological intervention measures, this method breaks through the limitations of high labor costs, high economy, and difficulties in implementing personalized plans, achieving efficient and innovative breakthroughs. Clinical Trial: Chinese Clinical Trial Registry ChiCTR2500099927; <https://www.chictr.org.cn/bin/project/edit?pid=256631ChatGPT?Total> Knee Arthroplasty?Perioperative?Anxiety?Depression.

(JMIR Preprints 06/04/2025:75557)

DOI: <https://doi.org/10.2196/preprints.75557>

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 a JMIR publication, my accepted manuscript PDF will be made available to all users.

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Impact of large language model Dialogue System on Perioperative Anxiety and

Depression in Patients Undergoing Total Knee Arthroplasty: A Randomized Controlled Trial

Yi-xiang Zhang^{1,2#}, Chun-lei Wu^{2,3#}, Tao Yang², Hao-tian Yin^{1,4}, Jing-chao Du¹, Xin Fu² Jun Liu^{1,2*}.

1. Department of Joints, Tianjin Hospital, Tianjin, China.
2. Clinical College of Orthopedics, Tianjin Medical University, Tianjin, China.
3. The united logistics security force of 983 hospital, Tianjin, China.
4. Department of Joints, Tianjin Hospital of Tianjin University (Tianjin Hospital), Tianjin, China

Abstract

Background: Total knee arthroplasty (TKA) is a widely performed procedure for knee osteoarthritis, yet patient satisfaction remains suboptimal, with preoperative anxiety and depression contributing to poor postoperative outcomes. Large-scale language models (LLMs), such as ChatGPT, have shown potential in healthcare applications but remain underexplored in the context of TKA. This study evaluates the impact of an AI-driven dialogue system on perioperative anxiety and depression in TKA patients.

Objectives—This study aimed to evaluate the effect of LLMs dialogue system on improving anxiety and depression, as well as overall psychological well-being, in patients undergoing TKA. The primary objective was to assess the system's impact on reducing perioperative anxiety and depression levels before and after surgery. Secondary objectives included evaluating patient satisfaction, the feasibility and scalability of the AI system compared to traditional methods, and its potential for broader clinical application in surgical settings.

Methods: We expect to conduct a parallel randomized and controlled trial in Tianjin Hospital. This trial content needs to gather 210 participants between the ages of 50-80 and randomly divide them into two groups, namely, the LLM intervention group or the usual care group. ChatGPT 4.0 software was used to conduct a personalized psychological education activity for about 20-30 minutes for the LLM intervention group every day, and the routine nursing group was used as the control group, receiving standard preoperative education. The test results were evaluated by the hospital anxiety and Depression Scale (HADS) at baseline, 48 hours, 24 hours before surgery, and 24 hours, 48 hours, 72

hours after surgery to determine whether artificial intelligence is effective in anti anxiety and depression, and improve patients' surgical mood and health.

Results: Statistical analyses showed that compared with the usual care group, the reduction of anxiety and depression in the LLM group was significantly greater. We found that both groups of patients experienced significant anxiety and depression before surgery. With the intervention and guidance of artificial intelligence, these symptoms in the LLM group continued to alleviate after surgery. The effect was particularly significant at 48 and 72 hours after surgery ($P < .001$), while the conventional nursing group showed a slower decline.

Conclusions: Experimental data shows that an intelligent psychological support system driven by artificial intelligence can effectively reduce anxiety and depression in TKA patients before and after surgery. Compared with traditional psychological intervention measures, this method breaks through the limitations of high labor costs, high economy, and difficulties in implementing personalized plans, achieving efficient and innovative breakthroughs.

Keywords □ Large-scale language models; ChatGPT; Total Knee Arthroplasty; Perioperative; Anxiety; Depression.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500099927; <https://www.chictr.org.cn/bin/project/edit?pid=256631>

Introduction

Knee osteoarthritis (OA) is a common musculoskeletal disease, and severe symptoms can lead to partial functional impairment, significantly reducing the quality of life of patients.¹ The existing total knee arthroplasty (TKA) is the mainstream surgery for the treatment of advanced severe knee osteoarthritis patients, with mature and efficient methods.² With the increasing prevalence of knee osteoarthritis worldwide, the demand for TKA surgery continues to rise. According to the 2021 report from the American Joint Replacement Registry (AJRR), the total number of TKA surgeries performed in the United States between 2012 and 2020 was approximately 12232993.³ In 2019,

China also recorded approximately 374833 cases, with a peak annual growth rate (CAGR) of 27.4%.⁴

In the past 40 years, TKA has become a very routine and widely performed surgery, but as far as we know, only about 80% of patients express complete satisfaction with the results. That data is not ideal.⁵⁻⁸ Despite significant advances in surgical techniques and patient management, approximately one in five TKA patients remain dissatisfied with the results.⁵ This dissatisfaction arises from multiple factors, including preoperative, intraoperative, and postoperative influences.⁹ Mental health issues, in particular, have gained increasing attention as a critical preoperative determinant. Studies have shown a strong link between preoperative anxiety or poor mental health and postoperative dissatisfaction.^{10,11} Specifically, preoperative anxiety and depression not only worsen postoperative pain but also increase the risk of insomnia and other adverse effects.^{12,13} For example, Chen et al. found that patients with preoperative psychological symptoms were more likely to experience chronic or worsening pain post-surgery.¹⁴ Gregory S. et al. also identified that psychosocial factors, such as anxiety, depression, kinesiophobia, central sensitization, and pain catastrophizing, are often linked to poor outcomes following TKA. As a result, addressing preoperative anxiety and depression has become a major challenge for joint surgeons.¹⁵

Large-scale language models (LLMs) are artificial intelligence (AI) systems trained on vast amounts of data from diverse sources, such as articles, books, and online content.^{16,17} These models learn the relationships between words and apply these patterns to various natural language processing (NLP) tasks. LLMs are artificial intelligence (AI) systems trained on large amounts of data from different sources such as articles, books, and online content. These models learn the relationships between words and apply these patterns to various natural language processing (NLP) tasks. LLM can also help people solve problems in their daily lives, such as translation, text synthesis, and sentiment analysis.¹⁷ With the influence of OpenAI's GPT series, especially the launch of ChatGPT in 2022, the development of large-scale language models has become a core

technological direction in the field of artificial intelligence, and has also received widespread attention for its ability to help humans synthesize text.¹⁸ The recent release of GPT-4 has pushed this capability to a climax, as ChatGPT not only handles multimodal inputs such as images, text, and audio, but also expands its application scope.¹⁹ So far, ChatGPT has achieved some success in medical applications, such as assisting in diagnosis, summarizing medical literature, and providing patient education.²⁰⁻²² However, its use in joint surgery, particularly for patients undergoing TKA, remains limited.

This study aimed to assess the effectiveness of a large language model dialogue system in delivering psychoeducation and psychotherapy over six consecutive perioperative days to reduce anxiety and depression in patients undergoing TKA with regional anesthesia and sedation. If proven feasible, this approach could offer a novel strategy for reducing surgical anxiety, improving patient management, and enhancing satisfaction.

Methods

Research design and study sample

We conducted a parallel, randomized, controlled, single-center trial with two groups: one using a large language model dialogue system and the other receiving usual care. Participants were recruited from the Joint Surgery Center at Tianjin Hospital and had to meet the following inclusion criteria: aged 50–80 years and willingness to participate in perioperative intervention studies. Patients were excluded if they met any of the following criteria: 1) contraindications to TKA, such as local or systemic infections; 2) preparation for bilateral TKA or revision surgery; 3) severe cognitive impairment or psychiatric disorders; 4) verbal communication impairments or inability to use a smart device; or 5) prior treatment with other systemic psychological interventions. Additional exclusion criteria included sleep apnea, delirium, inflammatory diseases, age over 80 years, and the need for postoperative discharge to inpatient rehabilitation or skilled nursing facilities.

The study was approved by the Institutional Review Board (IRB) of Tianjin Hospital (IRB 2024

Medical Ethics Review 203) and was registered with ClinicalTrials.gov. Written informed consent was obtained from all participants prior to enrollment.

Recruitment

Participants were recruited from the orthopedic outpatient clinic at Tianjin Hospital between July and October 2024. Patients scheduled for initial unilateral TKA due to primary osteoarthritis, diagnosed by two experienced joint surgeons, were approached by the study coordinator during their clinic visits to assess their eligibility and willingness to participate. Written informed consent was obtained from all participants. Prior to the trial, the researcher generated randomized lists using a computer, which were then sealed in opaque envelopes for each group. During data collection, the researcher opened the envelopes and randomly assigned participants to either the LLM group or the usual care group. Data were collected at the following time points: at admission (baseline, T1), 48 hours preoperatively (T2), 24 hours preoperatively (T3), 24 hours postoperatively (T4), 48 hours postoperatively (T5), and 72 hours postoperatively (T6).

Intervention

Patients in the usual care group received standard preoperative education and perioperative care, including surgery-related education and psychological counseling provided by doctors and nurses. Patients in the experimental group received emotion- and anxiety-relieving dialogues assisted by the LLM Dialogue System (ChatGPT 4.0) for 20 to 30 minutes daily. ChatGPT 4.0 provides personalized counseling and pre-surgical education based on the patient's emotional feedback, including positive thinking exercises,²³ anxiety management skills,²⁴ and breathing exercises.^{25,26} The language model also answers patients' questions about the surgery, such as anesthesia methods, surgery duration, and pain management during recovery. The intervention period for both groups lasted from 72 hours preoperatively to 72 hours postoperatively, totaling six days. Patients underwent elective unilateral TKA under regional anesthesia. All TKAs were performed using a medial parapatellar approach with a posterior-stabilized prosthesis. For patients with severe patellar articular

surface damage, the surface was trimmed as needed. All prostheses were fixed with bone cement, and there were no differences in the surgical approach between the two groups.

Study measures

Demographic data, including age, gender, BMI, years of education, smoking history, alcohol consumption, and marital status, were collected and analyzed using a baseline questionnaire. The primary outcomes were anxiety and depression, assessed using the Hospital Anxiety and Depression Scale (HADS).²⁷ Both the anxiety and depression subscales consist of 7 items, each rated on a 4-point scale from 0 (not at all) to 3 (very much). Higher scores indicate more severe anxiety or depression. A subscale score of 0–7 was considered normal, 8–10 indicated a suspected case of anxiety or depression, and ≥ 11 indicated a definite case. The HADS is a reliable tool widely used to assess anxiety and depression in perioperative patients.²⁸ We hypothesized that the LLM group would significantly reduce patients' anxiety and depression compared to the usual care group. In this study, patient satisfaction was used as a secondary outcome indicator and the Patient Satisfaction Questionnaire (PSQ) was used to assess the satisfaction of patients in both groups in terms of treatment experience, pain management, doctor-patient communication and postoperative recovery.²⁹ Scores ranged from 1 to 5 (1 = very dissatisfied, 5 = very satisfied).

Statistical analysis

Statistical analyses were conducted using IBM SPSS Statistics for Windows (Version 27). Continuous variables were summarized as means and standard deviations or as medians and interquartile ranges, depending on the data distribution. Normality was assessed using the Shapiro-Wilk test, along with visual inspection of Q-Q plots and histograms. Categorical variables are presented as counts and percentages. Depending on data distribution, t-tests or Wilcoxon rank-sum tests were used to compare continuous outcomes between the large language model and usual care groups at individual time points. Regression analyses employing generalized estimating equations (GEE) and autoregressive [AR(1)] correlation structures compared outcomes measured at multiple

time points and adjusted for baseline measurements when appropriate. The GEE method accounts for correlations between repeated measurements from the same patient, whereas the AR(1) structure assumes higher correlation for measurements taken closer together in time. Initially, interaction terms between randomized groups and time points were included in each outcome model. Group-by-time interactions were included in the GEE model to evaluate treatment effect trends over time. If no interaction was detected, the model was refitted without the interaction term. If an interaction was detected, a separate effect test was performed, and results were reported as the overall perioperative difference within or between groups, with a 95% confidence interval. GEE analyses were also conducted to evaluate the trajectory of outcomes over time for both the large language model and usual care groups. All statistical tests were two-tailed, with $P < .05$ indicating statistical significance.

Results

Participant characteristics

Participants were recruited between June 2024 and October 2024. As shown in Fig. 1, 235 eligible patients were referred by two doctors from the outpatient department for enrolment in the study. During face-to-face consultations, the study protocol and objectives were explained to the patients. 13 patients were excluded due to disinterest or other commitments, while 12 patients did not proceed to surgery. Reasons for non-progression included: change of mind ($n=8$) and symptom improvement ($n=4$). As shown in Table 1, the sample included 210 patients (84 males and 126 females), aged 55-79 years, with a mean age of 69.24 years (SD 5.59 years). There were no statistically significant differences in baseline demographic data between the large language model dialogue group and the usual care group ($P=.906$) (Table 1).

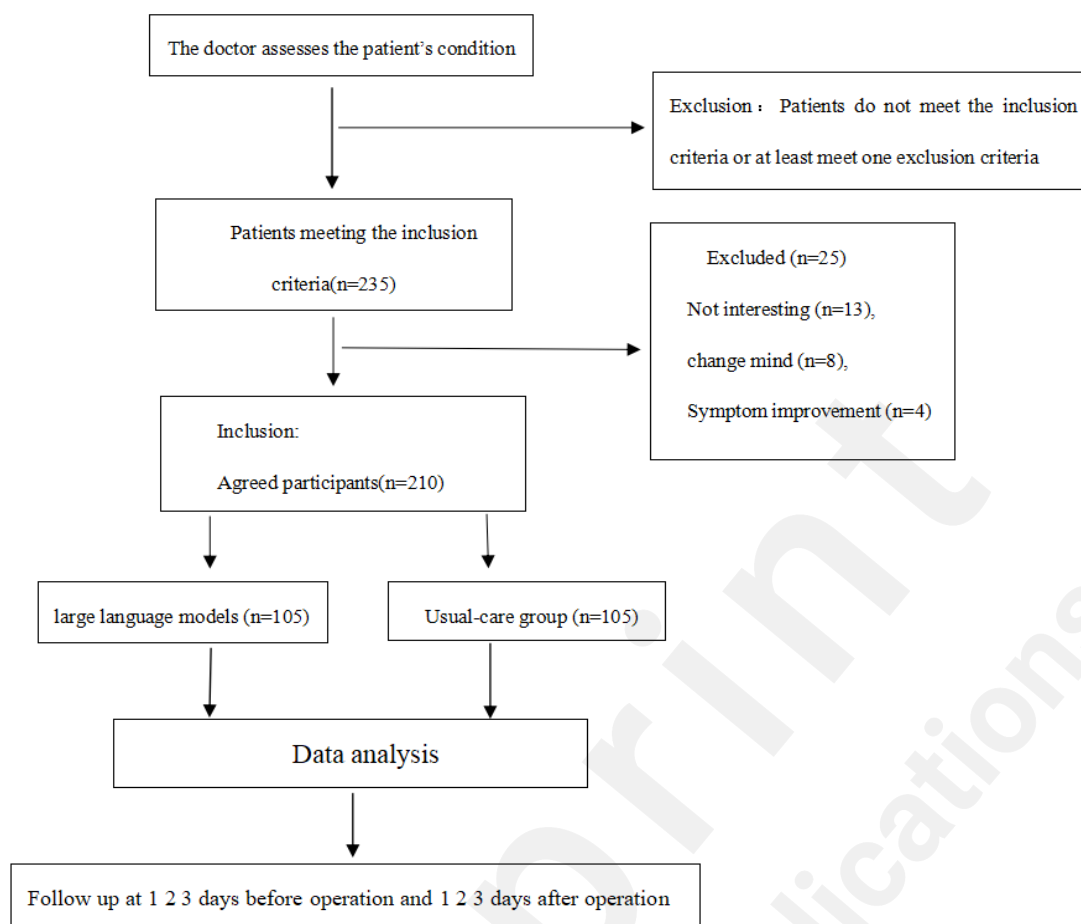


Figure 1 Consort Diagram of Screening, Randomisation, and Follow-up of Study Participants

Table 1 Characteristics of the Study Sample

Variables		Large language models (n=105)	Usual-care group (n=105)	P-value
Age, years: mean (SD)		69.1[6.0]	69.4[5.3]	0.906
Sex	Men	21(42%)	19(38%)	40(40%)
	Women	28(58%)	31(62%)	60(60%)
Education(years): mean (SD)		5.8[2.1]	5.6[1.9]	0.392
Body Mass Index, kg/m ² : mean (SD)		25.9[2.9]	26.7[3.3]	0.270
Surgical time (min): mean (SD)		93.3(2.5)	92.5(3.7)	0.697
Length of stay (d): mean (SD)		6.5(0.7)	6.8(1.2)	0.450
SD = Standard Deviation; BMI = Body Mass Index; Significance set at P value ≤ 0.05 . Continuous variables are reported as mean (standard deviation). Categorical variables are reported as n (%).				

Intervention effects on anxiety

Fig 2 illustrates the anxiety levels of the control and large language modelling groups at 72, 48,

and 24 hours preoperatively, and at 24, 48, and 72 hours postoperatively. At baseline, the mean anxiety score was 9.32 (SD 3.07), with no significant difference between the two groups at admission (usual care group: 8.95, large language model group: 9.70, $P = 0.258$). Generalized estimating equation (GEE) analyses revealed a significant interaction between group and time ($X^2_{\text{group*time}} = 432.31$, $p < .001$), prompting separate effects analyses. Table 2 summarizes the within-group test results for both groups at each time point, showing a statistically significant difference in anxiety between the groups at each time point ($p < .001$). Fig. 2 also demonstrates that both groups showed an increase in anxiety levels at 48 and 24 hours preoperatively, with a more pronounced increase in the control group. However, at 48 hours postoperatively, the large language modelling group showed a sustained decrease in anxiety levels, with a more significant recovery compared to the control group, which exhibited a more gradual decrease. Table 3 shows that the difference in anxiety levels between the two groups was statistically significant at all time points ($p < .001$), except for baseline levels. Anxiety levels in the large language modelling group were consistently lower than those in the control group at all time points, with the greatest difference (3.86 points) occurring 24 hours before surgery, and continuing to decrease postoperatively.

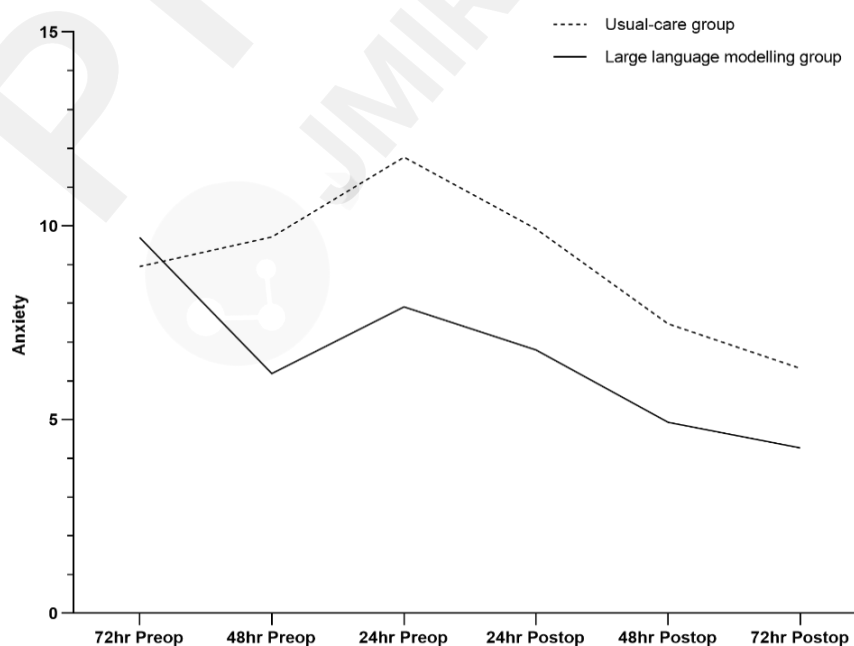


Figure 2 Changes in anxiety subscale scores across time according to a general linear model.

Table 2 Generalised linear model results on changes in anxiety outcomes over time in the Usual-care group and the large language modelling group

Timepoint		Estimate	SE	95% CI	P Value
Usual care group	48h Preop vs. Baseline	0.76	0.10	0.42–1.11	≤.001
	24h Preop vs. Baseline	2.82	0.13	2.37–3.27	≤.001
	24h Postop vs. Baseline	0.97	0.19	0.33–1.62	≤.001
	48h Postop vs. Baseline	-1.49	0.18	-2.11–-0.96	≤.001
	72h Postop vs. Baseline	-2.63	0.21	-3.33–-1.92	≤.001
Large language modelling groups	48h Preop vs. Baseline	-3.50	0.23	-4.26–-2.75	≤.001
	24h Preop vs. Baseline	-1.78	0.27	-2.67–-0.89	≤.001
	24h Postop vs. Baseline	-2.90	0.32	-3.96–-1.83	≤.001
	48h Postop vs. Baseline	-4.76	0.31	-5.79–-3.74	≤.001
	72h Postop vs. Baseline	-5.43	0.30	-6.44–-4.42	≤.001
Baseline each group is reference group. Difference in mean anxiety scores. Significance set at P value ≤ 0.05.					

Table 3 Generalised linear modelling results on the effect of large language models on anxiety

Group (LLM vs UC)		Estimate	SE	95% CI	P Value
Timepoint	Baseline	0.74	0.42	-0.67, 2.15	0.258
	48h Preop	-3.52	0.31	-4.56–-2.48	≤.001

	24h Preop	-3.86	0.36	-4.67□-3.05	□.001
	24h Postop	-3.12	0.22	-3.85□-2.40	□.001
	48h Postop	-2.53	0.18	-3.14□-1.92	□.001
	72h Postop	-2.06	0.16	-2.60□-1.51	□.001
Usual care group is reference group. Difference in mean anxiety scores. Significance set at P value □ 0.05.					

Intervention effects on depression

Fig 3 illustrates the depression levels in the Usual care and large language modelling groups at 72, 48, and 24 hours preoperatively, and at 24, 48, and 72 hours postoperatively. At baseline, the mean depression score was 4.42 (SD 1.18), with no significant difference between the two groups at admission (Usual care group: 4.56, large language modelling group: 4.28, $P = 0.258$). Generalized estimating equation (GEE) analyses revealed a significant interaction between group and time (Wald $X^2_{\text{group*time}} = 77.95$, $P < .001$), prompting separate effects analyses. Table 4 presents the within-group test results for both groups at each time point. The Usual care group showed significant differences at preoperative 48h, 24h, and postoperative 24h compared to baseline ($P < .001$), while the large language modelling group showed differences at preoperative 48h, postoperative 48h, and postoperative 72h compared to baseline ($P < .001$). Figure 3 shows that both groups exhibited an increase in depression levels at 48h and 24h preoperatively, similar to anxiety levels. However, the large language modelling group showed a significant decrease in depression levels postoperatively, particularly at 48h and 72h, compared to baseline. The Usual care group exhibited a more gradual decline. Table 5 shows that the difference in depression levels between the two groups was statistically significant at all time points ($P < .001$), except for the baseline level. Depression levels were lower in the large language modelling group than in the Usual care group at all time points,

with the greatest difference (1.64 points) at 48 hours postoperatively.

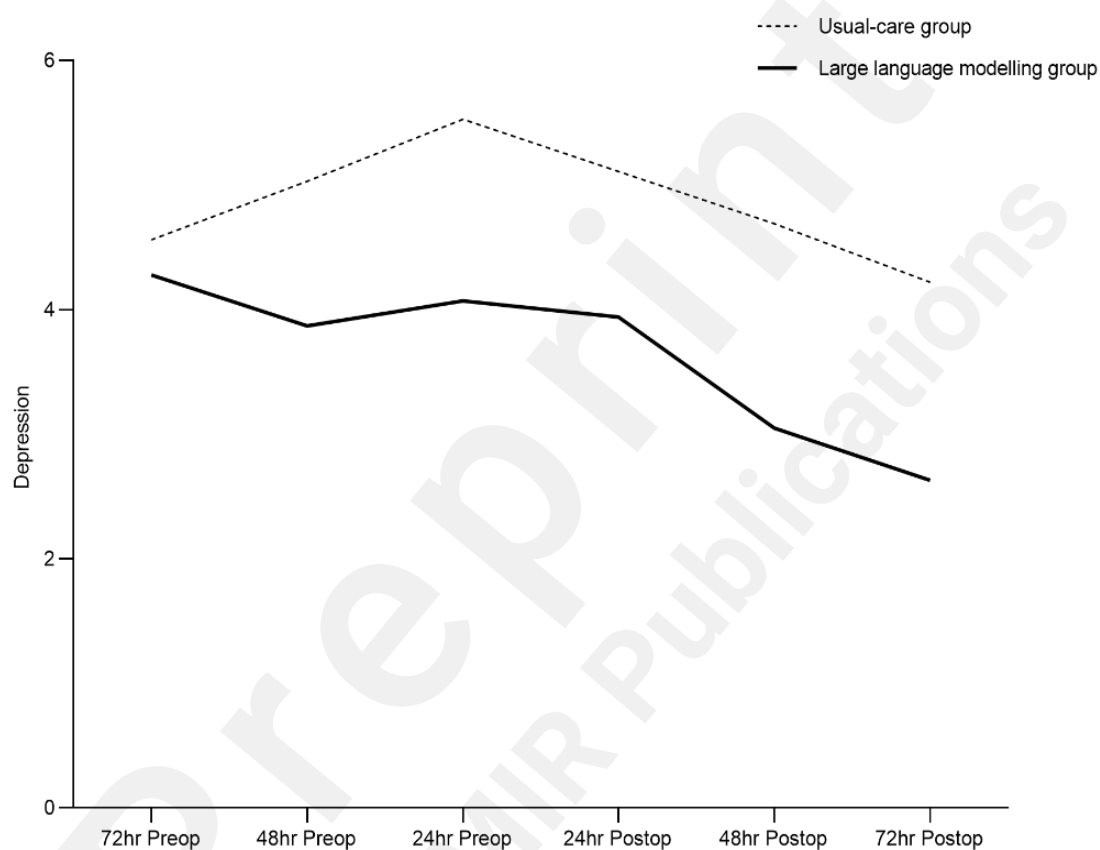


Figure 3 Changes in depression subscale scores across time according to a general linear model.

Table 4. Generalised linear model results on changes in depression outcomes over time in the Usual-care group and the large language modelling group

Timepoint		Estimate	SE	95% CI	P Value
Usual	48h Preop vs. Baseline	0.47	0.12	0.05–0.89	.010

care group	24h Preop vs. Baseline	0.97	0.13	0.54□1.41	□.001
	24h Postop vs. Baseline	0.97	0.19	0.33□1.62	□.001
	48h Postop vs. Baseline	0.12	0.17	-0.21,0.49	0.428
	72h Postop vs. Baseline	-0.34	0.13	-0.68,0.04	0.082
Large language modelling groups	48h Preop vs. Baseline	-0.41	0.08	-0.68□-0.13	□.001
	24h Preop vs. Baseline	-0.21	0.13	-0.59,0.15	0.248
	24h Postop vs. Baseline	-0.33	0.23	-0.77,0.13	0.162
	48h Postop vs. Baseline	-1.23	0.13	-1.65□-0.81	□.001
	72h Postop vs. Baseline	-1.65	0.12	-2.05□-1.24	□.001
Baseline each group is reference group. Difference in mean anxiety scores. Significance set at P value □ 0.05					

Table 5. Generalised linear modelling results on the effect of large language models on depression

Group (LLM vs UC)		Estimate	SE	95% CI	P Value
Timepoint	Baseline	-0.29	0.16	-0.75,0.19	0.240
	48h Preop	-1.16	0.15	-1.67□-0.66	□.001
	24h Preop	-1.47	0.23	-2.01□-0.93	□.001
	24h Postop	-1.17	0.21	-1.67□-0.74	□.001
	48h Postop	-1.64	0.97	-1.96□-1.31	□.001
	72h Postop	-1.59	0.89	-1.89□-1.25	□.001
Usual care group is reference group. Difference in mean anxiety scores. Significance set at P value □ 0.05.					

Patient Satisfaction Analysis

The LLM intervention group and usual care group had mean satisfaction scores of 4.1 (SD 0.7)

and 4.0 (SD 0.8), respectively ($P = 0.450$). Although the LLM group scored slightly higher in doctor-patient communication (4.3 vs. 4.1, $P = 0.120$), and in pain management and postoperative recovery, these differences were not statistically significant ($P = .335$).

Overall, while the LLM intervention group showed slightly higher satisfaction in some areas, no significant difference in overall patient satisfaction was observed between the two groups. This suggests that AI-driven interventions may not substantially impact patient satisfaction in the short term, and further studies are needed to evaluate their broader applicability in clinical settings.

DISCUSSION

Introduction to the Study

This study is the first to investigate the psychological effects of an AI driven intelligent psychological support system on anxiety and depression in patients undergoing total knee arthroplasty (TKA) before and after surgery. Anxiety and depression are common preoperative emotions in TKA patients, often exacerbated by fear, pain, and surgical success rates.³⁰ These negative psychological states can directly affect the postoperative pain control effect, delay the rehabilitation process, and reduce patients' satisfaction with the surgical outcome.^{5,9,31} According to our investigation, both anxiety and depression significantly increased before surgery, which is consistent with previous research.

AI-Driven Dialogue Systems and Their Effectiveness

Through comparative experiments, the LLM experimental group had more ideal results compared to the conventional treatment group, and the anxiety and depression of the former group patients were significantly reduced before and after surgery. This also provides a promising alternative to psychological support in surgical environments - artificial intelligence systems. The research results show that using AI driven dialogue systems can significantly alleviate anxiety and depression during

surgery. This conclusion indicates that preoperative emotional support can have a lasting impact on patients' mental health and postoperative recovery.

Advantages of AI Systems over Traditional Methods

Compared to traditional methods, systems driven by AI have multiple advantages. Usually the traditional preoperative guidance is generally provided by medical staff according to the unified template, which is more systematic, old-fashioned and difficult to meet the special psychological needs of each patient. Relatively speaking, AI driven dialogue systems provide real-time and personalized support, such as anxiety management methods, relaxation exercises, and tailored emotional interventions based on specific moods. This one-on-one intelligent counseling method helps alleviate anxiety and depression in different types of patients, creating a good atmosphere for their physical and mental health before and after surgery.

Scalability and Accessibility of AI-Driven Systems

In addition to personalized services, AI driven systems have scalability and accessibility. Unlike traditional manual intervention measures, it can simultaneously serve a large number of patients (without the limitations of hospital manpower and time - scalability), and its cost is lower, which makes it cheaper to provide intelligent psychological dialogue in remote or economically underdeveloped areas. No need for professional personnel to monitor the entire process, suitable for medical promotion under different conditions, which provides an undeniable advantage for its practical application.

Limitations of the Study

However, some issues were discovered during the research process: Firstly, patients with lower education levels or unfamiliar with modern intelligent networks, as well as the majority of elderly people, may not be able to operate and require the assistance of clinical doctors or nurses. Secondly, The duration of the intervention in this trial is limited to 6 days, and further follow-up research is needed to determine whether it can continue to improve the psychological health status of patients

during the postoperative recovery period.

Future Directions and Recommendations

Another limitation that cannot be ignored in this study is the subjectivity of the dialogue. In other words, the psychological questionnaire filled out by patients may be influenced by "response concerns", such as deliberately concealing their true feelings. In the future, objective measurements of emotional distress can be involved, such as wearing wristbands or monitoring physiological indicators such as heart rate and blood pressure through wearable devices, to supplement subjective questionnaire results and provide a more comprehensive assessment.

Conclusions

This randomized controlled trial demonstrates that an AI-driven dialogue system (ChatGPT 4.0) significantly reduces perioperative anxiety and depression in patients undergoing total knee arthroplasty (TKA). Compared to usual care, the intervention group exhibited a more pronounced and sustained reduction in anxiety and depression scores, particularly at 48 and 72 hours postoperatively ($P < 0.01$). The AI system's ability to deliver personalized psychological support, including anxiety management techniques, breathing exercises, and tailored surgical education, addresses the limitations of traditional standardized interventions, which often fail to meet individual patient needs.

Acknowledgements

The authors would like to sincerely thank the volunteers for their participation, which made this study possible!

Competing interests

The authors declare no competing interests.

List of abbreviations

TKA Total knee arthroplasty

LLMs Large-scale language models

HADS Hospital anxiety and Depression Scale

OA Osteoarthritis

AI Artificial intelligence

NLP Natural language processing

AJRR American Joint Replacement Registry

PSQ Patient Satisfaction Questionnaire

GEE Generalized estimating equations

AR Autoregressive

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

References

1. Glyn-Jones S, Palmer AJR, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. *The Lancet*. 2015;386(9991):376-87.
2. Carr AJ, Robertsson O, Graves S, Price AJ, Arden NK, Judge A, et al. Knee replacement. *The Lancet*. 2012;379(9823):1331-40.
3. Siddiqi A, Levine BR, Springer BD. Highlights of the 2021 American Joint Replacement Registry Annual Report. *Arthroplasty Today*. 2022;13:205-7.
4. Huang L, Wang Q, Zhao X. A commentary on "Incidence, patterns and risk factors for readmission following knee arthroplasty in China: A national retrospective cohort study" (*Int. J. Surg.* 2022;104:106759). *International Journal of Surgery*. 2022;106:106875.
5. Nam D, Nunley RM, Barrack RL. Patient dissatisfaction following total knee replacement. *The Bone & Joint Journal*. 2014;96-B(11_Supple_A):96-100.
6. Von Keudell A, Sodha S, Collins J, Minas T, Fitz W, Gomoll AH. Patient satisfaction after primary total and unicompartmental knee arthroplasty: An age-dependent analysis. *The Knee*. 2014;21(1):180-4.
7. Clement ND, Burnett R. Patient satisfaction after total knee arthroplasty is affected by their general physical well-being. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2013;21(11):2638-46.
8. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ. Patient Satisfaction after Total Knee Arthroplasty: Who is Satisfied and Who is Not? *Clinical Orthopaedics and Related Research*®. 2010;468(1).
9. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient Dissatisfaction Following Total Knee Arthroplasty: A Systematic Review of the Literature. *The Journal of Arthroplasty*. 2017;32(12):3854-60.
10. Scott CEH, Howie CR, MacDonald D, Biant LC. Predicting dissatisfaction following total knee replacement. *The Journal of Bone & Joint Surgery British Volume*. 2010;92-B(9):1253-8.
11. Huijbregts HJTAM, Khan RJK, Fick DP, Jarrett OM, Haebich S. Prosthetic alignment after total knee replacement is not associated with dissatisfaction or change in Oxford Knee

- Score: A multivariable regression analysis. *The Knee*. 2016;23(3):535-9.
12. Brander V, Gondek S, Martin E, Stulberg SD. THE JOHN INSALL AWARD: Pain and Depression Influence Outcome 5 Years after Knee Replacement Surgery. *Clinical Orthopaedics and Related Research®*. 2007;464.
 13. Pivec R, Issa K, Given K, Harwin SF, Greene KA, Hitt KD, et al. A Prospective, Longitudinal Study of Patient Satisfaction Following Total Knee Arthroplasty Using the Short-Form 36 (SF-36) Survey Stratified by Various Demographic and Comorbid Factors. *The Journal of Arthroplasty*. 2015;30(3):374-8.
 14. Chen D, Yang H, Yang L, Tang Y, Zeng H, He J, et al. Preoperative psychological symptoms and chronic postsurgical pain: analysis of the prospective China Surgery and Anaesthesia Cohort study. *British Journal of Anaesthesia*. 2024;132(2):359-71.
 15. Kazarian GS, Anthony CA, Lawrie CM, Barrack RL. The Impact of Psychological Factors and Their Treatment on the Results of Total Knee Arthroplasty. *JBJS*. 2021;103(18).
 16. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nature Medicine*. 2023;29(8):1930-40.
 17. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large Language Models in Medicine: The Potentials and Pitfalls. *Annals of Internal Medicine*. 2024;177(2):210-20.
 18. Huh J, Park S, Lee JE, Ye JC. Improving Medical Speech-to-Text Accuracy using Vision-Language Pre-training Models. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(3):1692-703.
 19. Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, et al. Summary of ChatGPT-Related research and perspective towards the future of large language models. *Meta-Radiology*. 2023;1(2):100017.
 20. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *New England Journal of Medicine*. 2023;388(13):1233-9.
 21. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*. 2022;23(6):bbac409.
 22. Singh S, Djalilian A, Ali MJ. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Seminars in Ophthalmology*. 2023;38(5):503-7.
 23. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. *Clinical Psychology Review*. 2018;59:52-60.
 24. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation Programs for Psychological Stress and Well-being: A Systematic Review and Meta-analysis. *JAMA Internal Medicine*. 2014;174(3):357-68.
 25. Balban MY, Neri E, Kogon MM, Weed L, Nouriani B, Jo B, et al. Brief structured respiration practices enhance mood and reduce physiological arousal. *Cell Reports Medicine*. 2023;4(1).
 26. Johnson MA, Simonian N, Reggente N. Lightening the mind with audiovisual stimulation as an accessible alternative to breath-focused meditation for mood and cognitive enhancement. *Scientific Reports*. 2024;14(1):25553.
 27. Snaith RP. The Hospital Anxiety And Depression Scale. *Health and Quality of Life Outcomes*. 2003;1(1):29.
 28. Feinstein MB, Krebs P, Coups EJ, Park BJ, Steingart RM, Burkhalter J, et al. Current Dyspnea Among Long-Term Survivors of Early-Stage Non-small Cell Lung Cancer. *Journal of Thoracic Oncology*. 2010;5(8):1221-6.
 29. Thayaparan AJ, Mahdi E. The Patient Satisfaction Questionnaire Short Form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. *Medical Education Online*. 2013;18(1):21747.
 30. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age and Ageing*. 2016;45(2):228-35.
 31. Healey EL, Mallen CD, Chew-Graham CA, Nicholls E, Lewis M, Lawton SA, et al. Integrating case-finding and initial management for osteoarthritis, anxiety, and

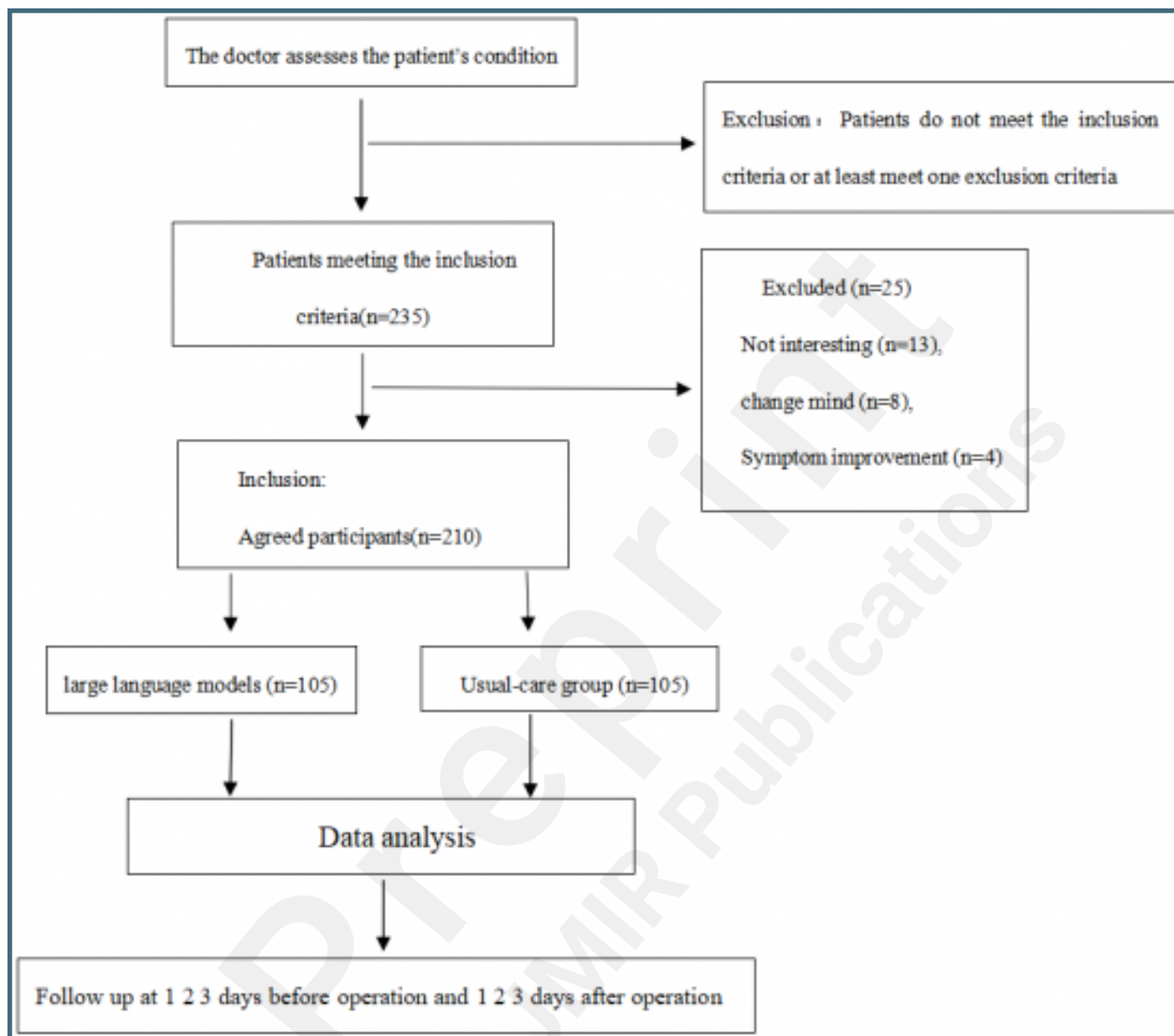
depression into primary care long-term condition reviews: results from the ENHANCE pilot trial. *Family Practice*. 2022;39(4):592-602.



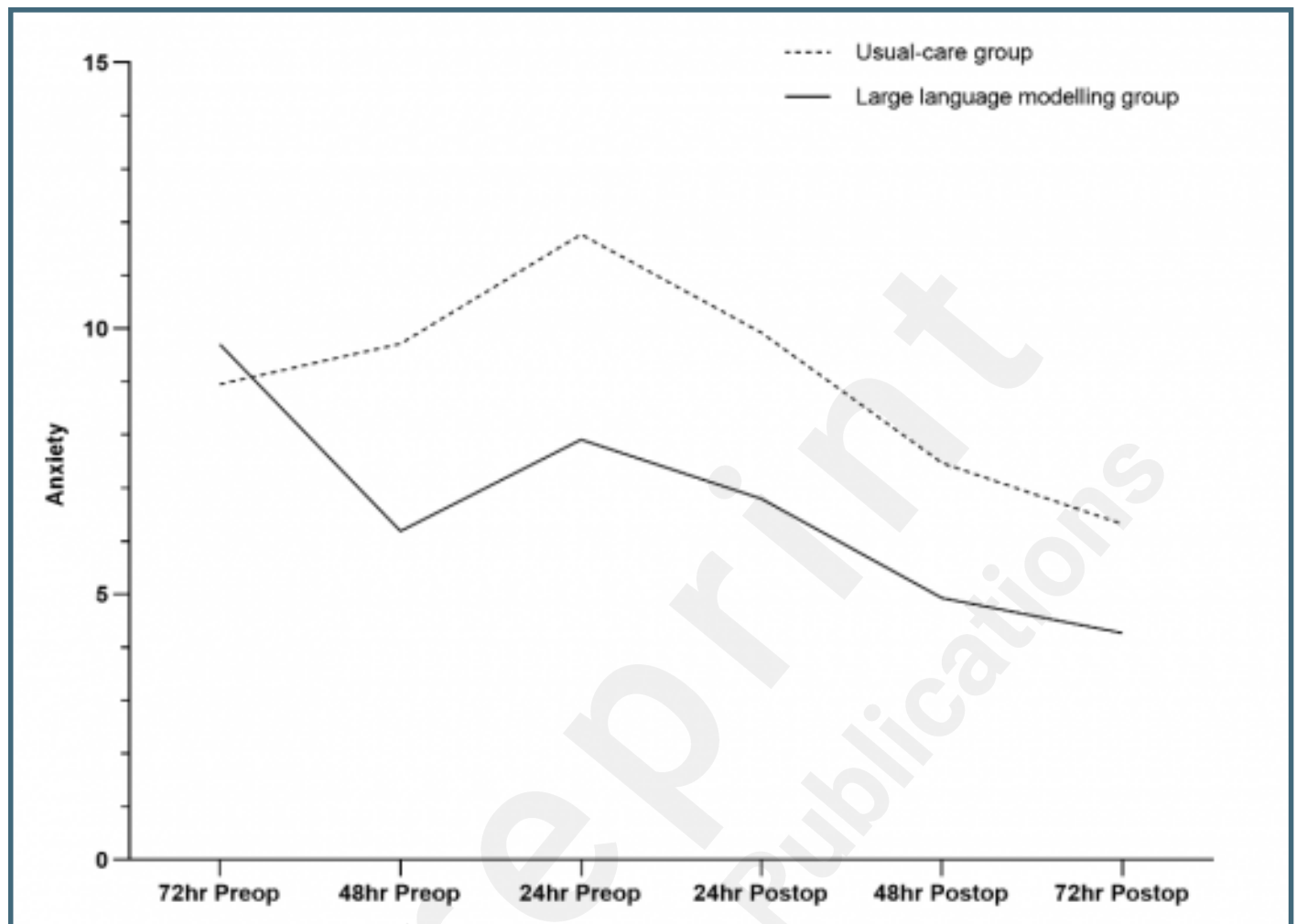
Supplementary Files

Figures

Consort Diagram of Screening, Randomisation, and Follow-up of Study Participants.



Changes in anxiety subscale scores across time according to a general linear model.



Changes in depression subscale scores across time according to a general linear model.

