Title: Artificial Intelligence in Depression – Medication Enhancement (AID-ME): A Cluster Randomized Trial of a Deep Learning Enabled Clinical Decision Support System for Personalized Depression Treatment Selection and Management

Authors: David Benrimoh*[1,2,3], Kate Whitmore[1], Maud Richard[1], Grace Golden[1,4], Kelly Perlman[1,2,3], Sara Jalali[3], Timothy Friesen[3], Youcef Barkat[3], Joseph Mehltretter[1], Robert Fratila[1], Caitrin Armstrong[1], Sonia Israel[1], Christina Popescu[1], Jordan F. Karp[5], Sagar V. Parikh[6], Shirin Golchi[7], Erica EM Moodie[7], Junwei Shen[7], Anthony J. Gifuni[3], Manuela Ferrari[3], Mamta Sapra[8,9], Stefan Kloiber[10,11,12,13], Georges-F. Pinard[14], Boadie W. Dunlop[15], Karl Looper[2,20], Mohini Ranganathan[16,17], Martin Enault[18], Serge Beaulieu[2,3,19], Soham Rej[20,21,22], Fanny Hersson-Edery[23], Warren Steiner[2], Alexandra Anacleto[1], Sabrina Qassim[1,24], Rebecca McGuire-Snieckus[25], Howard C. Margolese[2]

Affiliations:

1. Aifred Health Inc., Canada, Quebec, Montreal, H3J 1M1
2. Department of Psychiatry, McGill University, Canada, Quebec, Montreal, H3A 0G4
3. Douglas Mental Health University Institute, McGill University, Canada, Quebec, Verdun, H4H 1R3
4. University of Western Ontario, Canada, ON, London, N6A 3K7
5. Department of Psychiatry, University of Arizona, United States, Arizona, Tucson, 85721
6. Department of Psychiatry, University of Michigan, United States, Michigan, Ann Arbour, 48109
7. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada, Quebec, Montreal, H3A 0G4
8. Department of Psychiatry, Salem Veteran Affairs Medical Center, United States, Virginia, Salem, 24153
9. Virginia Tech Carilion School of Medicine, United States, Virginia, Roanoke, 24016
10. Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Canada, Ontario, Toronto M5T 1R8
11. Department of Psychiatry, University of Toronto, Canada, Ontario, Toronto, M5S 1A1
12. Institute of Medical Science, University of Toronto, Canada, Ontario, Toronto, M5S 1A1
13. Department of Pharmacology and Toxicology, University of Toronto, Canada, Ontario, Toronto, M5S 1A1
14. Department of Psychiatry, Institut Universitaire en Santé Mentale de Montréal, Canada, Quebec, Montreal, H1N 3M5
15. Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, United States, Georgia, Atlanta, 30322
16. Yale University School of Medicine, United States, Connecticut, New Haven, 06510
17. Veterans Affairs Connecticut Healthcare System, United States, Connecticut, West Haven, 06516
18. Relief - The Path of Mental Health, Canada, Quebec, Montreal, H2L 1J6
19. Bipolar Disorders Clinic, Douglas Mental Health University Institute, McGill University, Canada, Quebec, Verdun, H4H 1R3
20. Department of Psychiatry, Jewish General Hospital, Canada, Quebec, Montreal, H3T 1E2
21. Lady Davis Institute, Jewish General Hospital, Canada, Quebec, Montreal, H3T 1E2
22. School of Nursing and Midwifery, Queen's University Belfast, United Kingdom, Northern Ireland, Belfast, BT9 7BL
23. Department of Family Medicine, McGill University, Canada, Quebec, Montreal H3A 0G4
24. University of Waterloo, Ontario, Quebec
25. Barts and the London School of Medicine, United Kingdom, England, London, E1 2AD

*Corresponding Author:

David Benrimoh, MD.CM., MSc., MSc., FRCPC
Douglas Mental Health University Institute
Phone: +1 (514) 463-7813
Email: david.benrimoh@mail.mcgill.com

Abstract:
Background: There has been increasing interest in the use of Artificial Intelligence (AI)-enabled clinical decision support systems (CDSS) for the personalization of major depressive disorder (MDD) treatment selection and management, but clinical studies are lacking. We tested whether a CDSS that combines an AI which predicts remission probabilities for individual antidepressants and a clinical algorithm based on treatment can improve MDD outcomes.
Methods: This was a multicenter, cluster randomized, patient-and-rater blinded and clinician-partially-blinded, active-controlled trial that recruited outpatient adults with moderate or greater severity MDD. All patients had access to a patient portal to complete questionnaires. Clinicians in the active group had access to the CDSS; clinicians in the active-control group received patient questionnaires; both groups received guideline training. Primary outcome was remission (<11 points on the Montgomery Asberg Depression Rating Scale (MADRS)) at study exit.
Results: 47 clinicians were recruited at 9 sites. Of 74 eligible patients, 61 patients completed a post-baseline MADRS and were analyzed. There were no differences in baseline MADRS ($p = 0.153$). There were more remitters in the active (n= 12, 28.6%) than in the active-control (0%) group ($p = 0.012$, Fisher's exact). Of three serious adverse events, none were caused by the CDSS. Speed of improvement was higher in the Active than the Control group (1.26 vs. 0.37, $p = 0.03$).
Conclusions: While limited by sample size and the lack of primary care clinicians, these results demonstrate preliminary evidence that longitudinal use of an AI-CDSS can improve outcomes in moderate and greater severity MDD.

Keywords:
Major depression, clinical decision support system, artificial intelligence, machine learning

Introduction

Major depressive disorder (MDD) is a leading cause of disability and socioeconomic burden[1] impacting more than 300 million people worldwide[2]. Unfortunately, only a minority of patients will improve with the first treatment trial, and repeated treatment trials have diminishing probabilities of success[3]. Many patients undergo an arduous " trial and error" treatment selection approach, resulting in poorer outcomes, longer time in treatment, and greater patient and family burden[4]. To improve outcomes, it would be valuable to have a scalable point-of-care tool which can help personalize treatment choice without requiring expensive testing[5,6].

There have been several efforts in recent years to use artificial intelligence (AI) to predict treatment outcomes in order to better match patients to specific treatments (see[7]. Most studies have differentiated between two treatments (e.g. two drugs) or treatment types (e.g. two types of psychotherapy), limiting clinical utility when many treatments are available. In addition, clinicians are often concerned about model bias and being able to interpret the outputs of AI predictive models which are often considered to be "black boxes"[8,9-10] (providing predictions without clear reasoning). In addition, while improving predictions about treatment outcome may

be helpful to personalize treatment, previous work has shown that treatments are often not managed in accordance with guidelines in terms of dosage and monitoring[11–13]. There is a need for a solution to both the treatment *selection* and the treatment *management* problems while integrating into existing clinical workflows[14].

To address this, Aifred investigators developed the Aifred Clinical Decision Support System (hereinafter referred to as the CDSS). This is a digital platform which supports clinicians in the implementation of guidelines (2016 CANMAT depression guidelines[15]) and measurement-based care[15] in order to solve the treatment *management* problem, and which includes an AI (deep learning) powered module to assist in baseline treatment *selection* by providing predicted probabilities of remission for 10 commonly used first line antidepressants and combinations of these. Extensive feasibility and ease of use testing of this CDSS was previously performed in both simulation center and *in vivo* feasibility studies[16–19]. With *in silico* testing demonstrating that the AI component should help improve remission rates[6,20–22] and *in vivo* testing demonstrating that the platform was feasible, easy to use and likely safe[16–19] the current study was undertaken with the main objective of determining the efficacy of the platform in improving depression treatment outcomes in patients with moderate to severe depression, as well as to assess platform safety.

Methods:

This study is reported as per the CONSORT-AI checklist [23] and was conducted in accordance with all relevant ethical regulations including the Declaration of Helsinki and the Tri-Council Policy Statement. The research ethics board of the Douglas Research Center gave ethical approval for this work and it was subsequently approved by central and local ethics review boards for each site. Written informed consent was obtained from all study participants. Full methods are available in the supplementary material.

Design

The current study is a two-arm, cluster-randomized trial, with clinicians serving as the cluster. Clinicians rather than patients were randomized as they were the ones receiving the decision support intervention, and to avoid contamination [24]. Patients entered the intervention arm of their treating clinician.

Participants - Clinicians

Clinicians could include primary care doctors, psychiatrists, residents, nurse practitioners (with or without specialized mental health training), who saw at least one patient with depression per month, on average, before study start.

Participants - Patients

Patient recruitment criteria were intended to be broad in order to replicate a naturalistic outpatient depression population with moderate to severe depression. Patients were recruited from the practices or hospital-based clinics of the participating clinicians. Inclusion criteria were as follows: 1) age 18 and over 2) diagnosed by their treating clinician with MDD using DSM-5 criteria [25] 3) MDD diagnosis confirmed via a blinded rater who completed the Mini Neuropsychiatric Interview (MINI) [26] and 4) at least moderate severity, as assessed by a blinded rater completing the Montgomery Asberg Depression Rating Scale (using a cutoff of 20) [27]. Exclusion criteria were as follows: 1) age under age 18; 2) presence of bipolar disorder of any

type; 3) inability or unwillingness to give informed consent; 4) inability to manage patient safely as an outpatient; 5) an active major depressive disorder was not the main condition being treated; and 6) an inability to use the tool (e.g. because of severe cognitive impairment). An active major depression meant that the depression, in the judgment of the treating clinician, required an initiation or a change in treatment. Psychiatric comorbidities (aside from bipolar disorder) were permitted.

Settings

Eligible settings included any primary or secondary/tertiary public or private outpatient setting in the United States or Canada which provided outpatient care for patients presenting with MDD. A diverse array of sites joined the study. These included public sector psychiatric clinics in Canada and university-affiliated and Veteran's Affairs mental health services in the U.S.

Intervention - Aifred CDSS

The Aifred CDSS platform consists of the following elements. The *patient portal*, accessible by web browser or mobile phone application, allows patients to complete questionnaires, receive email reminders to complete questionnaires, visualize questionnaire scores and interpretations, and track treatments which they or their clinicians enter. The *clinician portal*, accessible by web browser, allows clinicians to see all the information patients enter, while at the same time giving them access to the *clinical algorithm* module. This module is a rule-based decision tree based on the CANMAT 2016 guidelines for depression treatment [15]. This module presents the clinician with patient-specific, guideline derived information about treatment options based on the patient's depression severity, change in depression severity over time (measured using the Patient Health Questionnaire (PHQ-9)[28]), and current treatments. The algorithm provides new information at each patient visit, based on patient progress. The AI component, which is separate from but housed within the algorithm module, is focused on assisting with treatment *selection* by generating remission probabilities for commonly used first line antidepressants based solely on baseline clinical and demographic details. The AI is a deep learning model trained on 9042 patients from depression treatment trials[29,30]; the AI, the data it intakes, and the report it provides clinicians is described further in the supplementary materials.

Intervention - Patients

All patients received access to the patient portal of the CDSS. Patients remained in the study for 12 weeks from their first treatment visit, at week 2, weeks 4-6, week 8 and week 12.

Intervention - Clinicians

There were two intervention groups: an Active group and an Active-Control group[31,32]. Clinicians in the Active-Control were provided with the results of questionnaires patients completed as well as training on the CANMAT guidelines[15]. Active group clinicians received guideline training, and were provided with full access to and training on the clinician portal of the Aifred CDSS. As all raw data provided to clinicians was the same in the Active and Active-Control groups, the only group differences consisted of the provision of the data processed by the clinical algorithm and AI model to the Active group. Clinicians were not required to use the information they were provided in any specific manner.

Measures

At baseline, the blinded rater assessed the Mini International Neuropsychiatric Interview (MINI) and Montgomery–Åsberg Depression Rating Scale (MADRS). Patients were asked to complete a Patient Health Questionnaire (PHQ-9) and General Anxiety Disorder (GAD-7) weekly once they had accounts on the CDSS. The MADRS was administered by the blinded rater at screening, visit 3 (weeks 4-6 of treatment), visit 4 (week 8) and visit 5 (week 12). Trained study staff also administered the Brief Adherence Rating Scale (BARS) after every visit to assess treatment adherence [33]. Further measures are detailed in the supplementary material.

Outcomes

The pre-specified primary outcome of the study was remission of depressive symptoms, defined as a score of <11 [34,35] on the MADRS at study exit for those patients with at least two MADRS scores. Safety outcomes included an examination of the nature and number of adverse and serious adverse events in each group. Secondary outcomes included response (50% reduction in symptoms) on the MADRS, rate of change of the MADRS score, and medication adherence using the BARS score.

Blinding

Patients

Patients were fully blinded to group assignment. They were told that they were entering a study where they would be using a new digital technology to provide information to their clinician and that there were two groups, but the nature of the groups was not revealed.

Clinicians

Clinicians were aware of their group assignments as they were the ones receiving the AI predictions. Clinicians were partially blinded to reduce expectation bias[36]: they were not told the study endpoints, and they were not informed of the expected effect sizes of the interventions.

Raters

Raters who collected the primary outcome (MADRS) and conducted the MINI were blind to group allocation.

Statistical Analysis

Outcome data were analyzed, as prespecified in the Statistical Analysis Plan, on an intent-to-treat basis for patients who had at least two ratings of the MADRS (the Analysis set). Safety data were analyzed for the Safety population, pre-specified as all patients who attended at least the first treatment visit. Demographic and baseline clinical data are presented in Table 1. The primary outcome (MADRS remission) was assessed using a Fisher's exact test. Secondary outcomes were compared using one-way ANOVAs, and proportions were compared using two-sided X^2 or two-sided Fisher's exact tests, as appropriate. Further pre-specified analyses are detailed in the Supplementary Material.

Early Study Termination

Unfortunately, due to lack of funding caused by delays related to the COVID-19 pandemic, the study was terminated early.

Results

Sites

Sites were located in Canada (5) and the United States (4) and included U.S.Veterans Affairs hospitals and mood disorders programs in university-affiliated psychiatric departments.

Recruitment - Clinicians

Of the 47 clinicians recruited who were cleared to recruit patients prior to early study termination, 25 were randomized to the Active group and 22 to the Active-Control group. 27 clinicians recruited at least one patient (57%); 16 in the Active group (64%) and 11 in the Active-Control group (50%).

Recruitment and Dropout - Patients

Patients were recruited between 2022-06-15 and 2023-11-16, a total of 17 months. The study was terminated early because of lack of funds due to delays in study initiation related to COVID-19. Recruitment and dropout are summarized in the CONSORT diagram **(Figure 1)**. Of the 74 eligible patients after screening and enrollment, 61 had at least 2 MADRS available, forming the Analysis set (n = 42 Active, n = 19 Active-Control). The groups did not differ in terms of 12 week completion (Active = 36/53 (68%); Active-Control = 18/21 (86%) (p = 0.15, Fisher's exact test)).

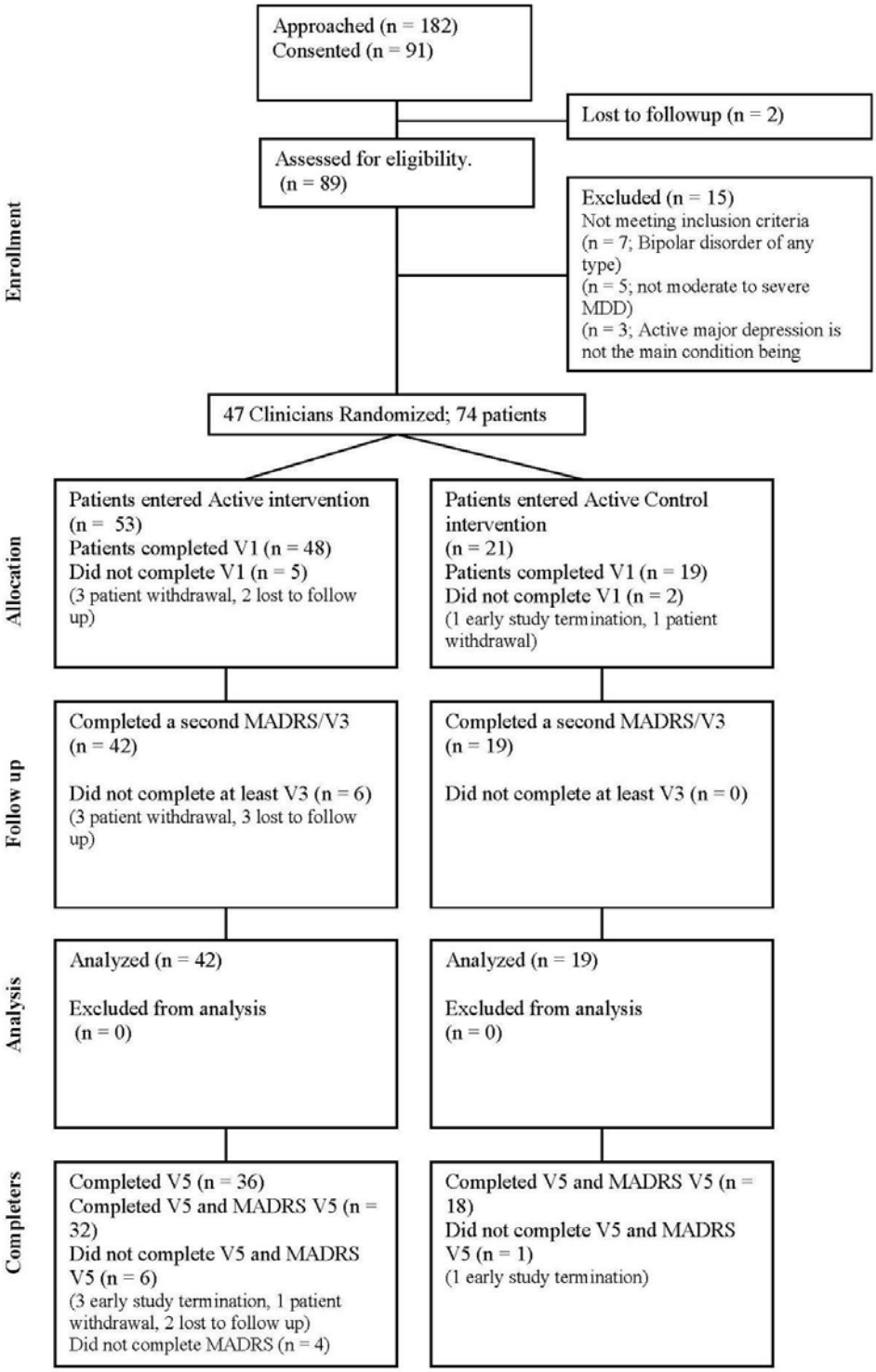


Figure 1. Patient Consort Diagram

Demographics and baseline clinical characteristics

Intervention and Active-Control groups did not differ with regards to important clinical and demographic characteristics (Table 1). The group had substantial chronicity of illness with 33 Active patients (78.6%) and 13 Active-Control patients (68.4%) having recurrent MDD. Further details are available in the Supplementary Material.

	Active group (n = 42)	Active-Control group (n = 19)
Mean age (SD)	44.0 (15.2)	39.3 (12.4)
Sex - n female (%)	20 (47.6)	10 (52.6)
Mean baseline MADRS (SD)	33 (7.3)	30 (5.8)
Race	n = 40	n = 18
- White	30 (75.0)	13 (72.2)
- Other**	10 (25.0)	5 (27.8)
Mean yearly household income (USD)	n = 37	n = 15
- Mean (SD)	42,206.87 (29103.3)	44,817.10 (17975.4)
Highest level of education achieved*	n = 40	n = 17
- Some high school/ high school diploma or equivalent (GED)	5 (15)	5 (27.8)
- Some university or college	12 (30.0)	4 (22.2)
- Bachelor's degree	11 (27.5)	6 (33.3)
- Graduate or Professional Degree	7 (17.5)	1 (5.6)
- Trade/technical training or other	5 (12.5)	1 (5.6)
Graduated high school	39 (97.5)	17 (94.4)
Currently employed	n = 39	n = 18
	18 (41.5)	12 (66.7)
Marital Status		
- Single	22 (55.0)	12 (66.7)

- Partnered	18 (45.0)	6 (33.3)
Mean number of medications (all indications) prescribed at baseline (n = 57)	3.18	3.06
Adverse Childhood Experiences (mean, (SD))	2.62 (2.59)	3.61 (2.59)
MINI Comorbidities: n (%)		
- Suicidality (current-past month)	22 (53.5)	13 (68.4)
- High suicidality score category	15 (35.7)	6 (31.6)
- Generalized anxiety disorder current	15 (35.7)	9 (47.4)
- Social anxiety disorder current	11 (26.2)	4 (21.1)
- Posttraumatic Stress Disorder current	10 (23.8)	1 (5.3)
- Panic Disorder current	7 (16.7)	2 (10.5)
- Alcohol Use Disorder Past 12 months	6 (14.3)	6 (31.6)
- Agoraphobia current	5 (12)	2 (10.5)
- Substance Use Disorder (Non-Alcohol), Past 12 months	5 (11.9)	2 (10.5)
SAPAS-SA (Personality disorder screening) N(%)	n = 41	n = 17
Those meeting cutoff score of 3 or more for positive screening	25 (61)	8 (47.1)

Table 1: Baseline Clinical and Demographic Characteristics per Group

*Note: participants could select more than one option; the graduated high school entry was constructed based on the available data.

** Lower count rows have been collapsed into the ‘other’ category in order to preserve confidentiality

Further demographic details are available in the Supplementary Materials.

Treatment Outcome - Remission at Study Exit

On the primary outcome, remission, there were significantly more remitters in the Active (n = 12/42 (28.6%)) than in the Active-Control (0/19, (0%)) group (p = 0.01, Fisher's exact test). Outcomes are summarized in table 2.

Treatment Outcome - Response and Change from Baseline

With respect to treatment response (defined as a 50% or greater decrease in total MADRS score between screening and study exit), 17 Active patients (40.5%) and 3 Active-Control patients (15.8%) responded at study exit. This was a large numerical difference, however, it did not reach significance (X^2 = 3.6, p = 0.06). The proportion of responders was not significantly different at visit 3 or 5, but was significantly different at visit 4 (p = 0.04, Fisher's exact test).

With respect to change from baseline to study exit, patients in the Active group experienced a mean 12.0 point improvement in MADRS score (SD = 13.5) while those in the Active-Control group experienced a change of 4.9 (SD = 10.9). Again, while a large numerical difference, it did not reach significance (F(1) = 4.006; p = 0.05). This corresponds to a between-group difference of 7.1 points, which exceeds the accepted threshold for a minimum clinically important difference between groups on the MADRS [37].

In terms of percent change of MADRS score from baseline to score at study exit, the Active group experienced a mean 35% change (SD = 41.1) and the Active-Control group experienced a mean 13.2% change (SD = 36.2); this difference was again numerically large but non-significant (F(1) = 3.95, p = 0.05).

Treatment outcome - Rate of Change

Investigators observed a significantly faster rate of improvement in MADRS score (change in MADRS score divided by treatment weeks a patient spent in the study) in the Active group compared to Active-Control. The mean change in total MADRS score per week in the Active group was 1.26 points (SD = 1.63); in the Active-Control group this was 0.37 points per week (SD = 0.91), (F(1) = 4.99; p = 0.03; eta-squared = 0.08, 95% CI [0,0.23], ANOVA). See Figure 2.

Outcome	Active	Active-Control	p-value
Remission	12 (28.6%)	0 (0%)	0.01
Response	17 (40.5%)	3 (15.8%)	0.06
Mean change from baseline	12 (SD = 13.5)	4.9 (SD = 10.9)	0.05
Percent change from baseline	35 (SD = 41.1)	13.2 (36.2)	0.05
Slope of improvement (amount of change per week)	1.26 (SD = 1.63)	0.37 (SD = 0.91)	0.03

Table 2: Summary of Outcomes

Treatment Adherence

Patients reported high levels of treatment adherence across visits in the study on the BARS questionnaire (mean 96.4% adherence, SD = 13 in Active; 95% adherence, SD = 10.9 Active-Control, F(1) = 0.76, p = 0.40).

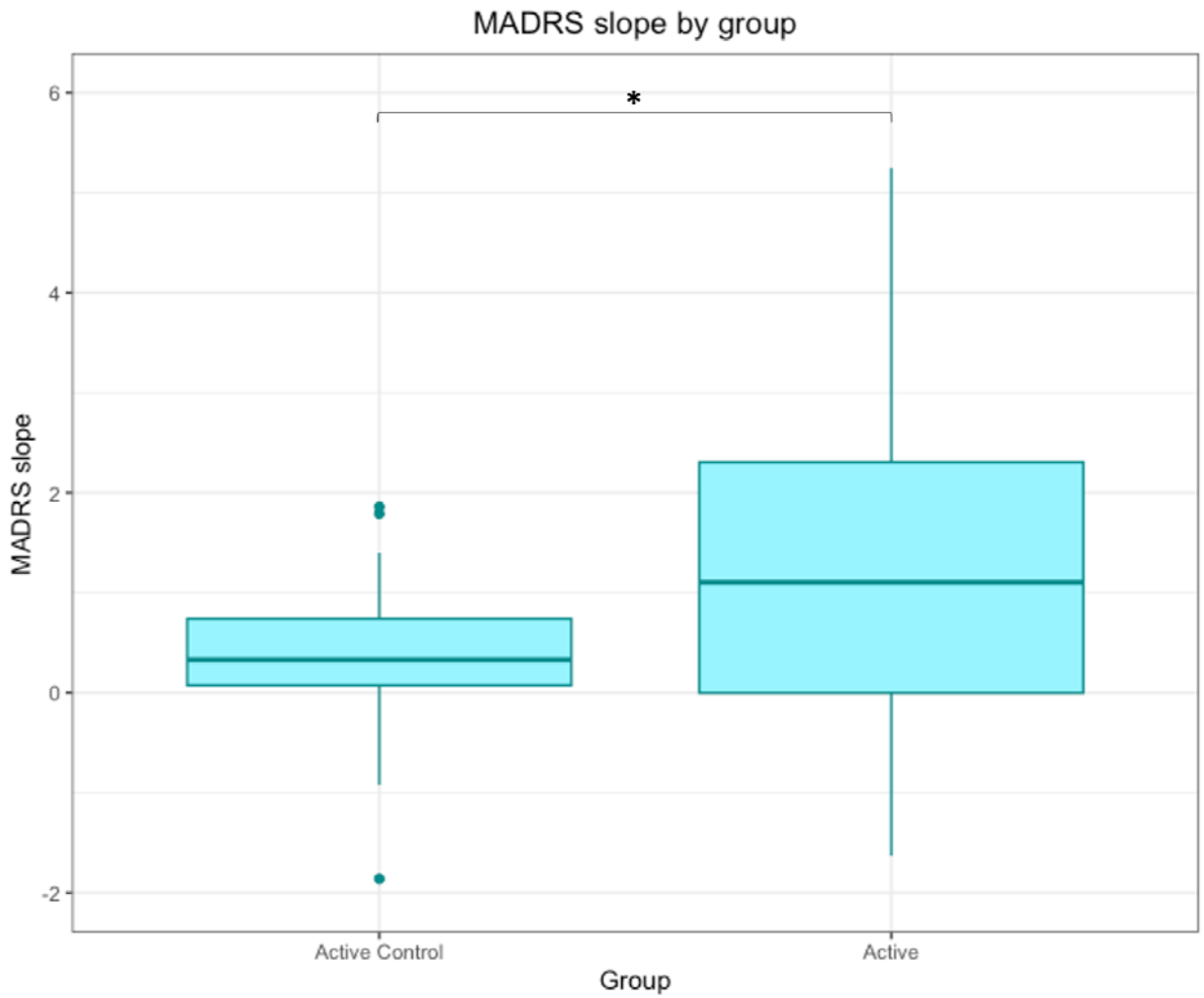


Figure 2: MADRS Slope by Group - bar plot displaying significantly different mean MADRS slope per group (F(1) = 4.99; p = 0.03; eta-squared = 0.08, 95% CI [0,0.23], ANOVA) Error bars represent standard deviation.

Safety Outcomes

With respect to safety, adverse event rates (e.g., medication side effects) and serious adverse event rates (e.g., hospitalizations) were examined in the Safety population (48 Active, 19 Active-Control). 89 adverse events were reported for the Active group, a rate of 1.9 events per patient, compared to 51 events (2.7 events/patient) in the Active-Control group. There were 3 serious adverse events in the Active group, and none in the Active-Control group. All three events were determined to have been unrelated to the CDSS by the site's primary investigator. Further details can be found in the Supplementary Material.

Patient Engagement

We examined PHQ-9 completion rates during the 12 treatment weeks (Fig. 3). The total PHQ-9 completion rate was 70% (67% in active, 77% in the Active-Control group). These completion rates are in line with our previous feasibility study[18,19].

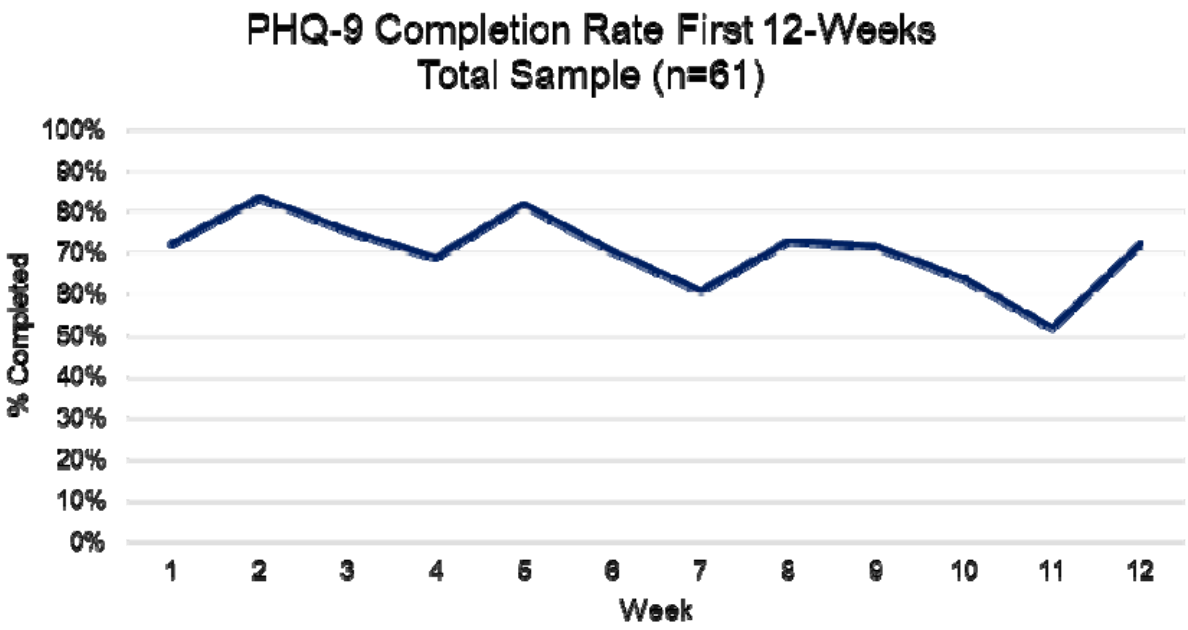


Figure 3: PHQ-9 completion rates for the 12 treatment weeks.

Clinician Engagement

Over the course of the study, 81.25% of doctors in the Active group who recruited patients accessed the CDSS at least twice (e.g. at least once after training); 81.25% accessed the treatment algorithm at least twice (e.g. at least once after training), and 81.25% accessed the AI results at least once (note: AI results were not available during training). Per visit, we report the number of Active clinicians who accessed the application, and of these, the number that accessed the treatment algorithm and the AI results (Table 3).

Visit	App Access (%)	Algorithm Access (%)	AI Result Access (%)
Treatment Visit 1 (Day 0)	0.88	0.69	0.67
Treatment Visit 2 (W2)	0.80	0.80	0.07

Treatment Visit 3 (W4-6)	0.74	0.74	0.00
Treatment Visit 4 (W8)	0.71	0.68	0.03
Treatment Visit 5 (W12)	0.56	0.47	0.00

Table 3: Clinician Platform Usage by Visit App access = proportion of clinicians in the active group who logged into the app at each visit; Algorithm Access = proportion of clinicians who went beyond logging in and access the clinical algorithm module; AI Result Access = proportion of clinicians who accessed the AI results at the end of one session of the clinical algorithm, which was intended to occur at Visit 1 or 2 (one access later in treatment was a result of an error made by a clinician and logged as a protocol deviation).

Discussion

This study is the first of its kind in mental healthcare to integrate an AI-powered CDSS in a longitudinal fashion to assist clinicians in making more effective clinical decisions about treatment selection and management while providing patients with more information about their own symptoms and trajectories. These initial results suggest that this CDSS is effective in improving remission rates (28.6% in the Active group compared to no remitters in the Active-Control group) and the rate of symptom improvement in adult patients with moderate or greater severity depression. In addition, engagement with the platform by both clinicians and patients was high throughout the study, and no adverse events were linked to the CDSS.

Despite both groups having the same baseline depression severity, similar clinical profiles, and the same treatment options permitted, patients in the Active group had significantly improved remission rates and more rapid improvement. Given this, we suggest that the CDSS may had a positive impact on clinical decision-making, demonstrating the potential for a CDSS which can organize information, present it at the clinically appropriate time at the point of care, and provide personalized treatment outcome predictions to significantly improve outcomes[38–40]. Further discussion of potential mechanisms of action for the CDSS is presented in the Supplementary Material.

This study has several strengths. The first is a design that intended to replicate realistic use of the CDSS, where clinicians and patients were not required to use the platform or adhere to it in any particular manner. Findings of improved outcomes in the Active group may therefore approximate benefits which would be derived in real clinical practice. The second keystrength is the comparison of the CDSS with an Active-Control group that approximated realistic best practices in-clinic today, suggesting that the CDSS may be able to improve outcomes over and above these best practices. Finally, the CDSS was simple to introduce into a diverse array of clinical environments and had reasoable patient and clinician engagement, which increases its potential for rapid adoption.

This study also has several limitations. The first is the smaller than intended sample size. This limits the power of subgroup and secondary analyses. It is reassuring, however, that significant clinical benefit was observed in line with the *a priori* estimated effect size. As discussed in the methods and in investigators' companion paper (Perlman et al. 2024), the AI model used in this study was limited to providing initial treatment outcome predictions and could not adapt to treatment failure, and it had a preference for escitalopram being predicted as the treatment most likely to be effective, while providing more variable predictions for the other medications (see[30] for detailed discussion). In addition, first line treatments such as

psychotherapy, vortioxetine and mirtazapine monotherapy could not be included in the AI model due to lack of data, though these were present in the treatment algorithm. Future versions of the model will continue to be improved by more and more diverse data, which will likely continue to improve the performance of the platform. Another limitation is the fact that the CDSS is a composite intervention, consisting of measurement based care, a rule-based algorithm, and the AI model. It would be important to be able to identify, in future work, which elements of the intervention are most responsible for the clinical improvements seen. Another limitation is the imbalance in the number of patients recruited into the Active and Active-Control groups. This may speak to the interest that Active group clinicians may have had in using the tool with patients.

Conclusions

In this paper, we have shown preliminary data indicating the clinical effectiveness and safety, in a cluster-randomized study, of an innovative AI-powered CDSS to support clinical decision making in the treatment of adult patients with MDD of moderate and greater severity. Use of this and similar systems, which could be implemented rapidly into clinical practice with minimal training, has significant potential to improve the effectiveness and speed of treatment for MDD. Future work on CDSS systems like this which are intended to be integrated into longitudinal care may benefit from the study methods discussed here. Replication of this work in larger samples by independent investigators is also needed. Future work could also be directed at further expanding and improving the AI model implemented in the intervention, potentially using data collected during real-world use of the CDSS.

Registration at clinicaltrials.gov: NCT04655924

The full protocol and statistical analysis plan are available in the Supplementary Material.

Funding: Funding was provided by Aifred Health; MEDTEQ FSISSS; Bell Let’s Talk and Brain Canada; Investissement Quebec; the Quebec Ministry of Economy and Innovation; the Mindstrong Foundation at the Jewish General Hospital; and the McGill Industry and Partnership grant. The study was co-designed and supervised by HM and Aifred Health. No other funder had a role in the development or reporting of this research.

DISCLOSURES

Declaration of Competing Interest:

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

DB, KW, MR, GG, KP, JM, RF, CA, SI, CP, AA, and SQ were employees and/or shareholders of Aifred Health Inc. and supported this research in the context of their work for Aifred Health. JFK, SP are members of Aifred Health's scientific advisory board. SP has received honoraria from Aifred Health. DB, SG, EEMM, and SR receive a salary award from the Fonds de recherche du Québec – Santé (FRQS). EEMM is a Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine. BWD has received research support from Boehringer Ingelheim, Compass Pathways, Intra-Cellular Therapies, NIMH, Otsuka, Usona Institute, and has served as a consultant for Biohaven, Cerebral Therapeutics, Myriad Neuroscience, and Otsuka. SR receives grant funding from Mitacs (for a graduate student), is on a steering committee for Abbvie, and owns shares of Aifred Health. SK reports grants from the Labatt Family Innovation Fund in Brain Health (Department of Psychiatry, University of Toronto), the Max Bell Foundation, the

Canadian Centre on Substance Use and Addiction, the Centre for Addiction and Mental Health Discovery Fund, the Ontario Ministry of Health and Long-Term Care (MOHLTC), the Canadian Institutes of Health Research (CIHR), and the International OCF Foundation (IOCDF). SK received honorarium for consultation from EmpowerPharm. JFK has been provided options in Aifred Health. HCM has received honoraria, sponsorship, or grants for his participation as a consultant, advisory committee member, and/or as a speaker at educational events for AbbVie, HLS Therapeutics Inc., Janssen, Lundbeck, Otsuka, Newron, Sunovion and Teva. He has received grants and/or research support from the MGH Hospital Foundation, SyneuRX and Aifred Health. SJ, TF, and YB recieved honoraria from Aifred Health for working on this paper. All other authors report no relevant conflicts.

Code availability: The methods used to develop the AI platform are available in previous publications; however the final code for the CDSS and final trained model are not publicly available at this time. Previous versions of the pipeline used to construct the model are available here (https://github.com/Aifred-Health/VulcanAI) and a full version of the model trained using a pharmaceutical dataset is available here: (https://github.com/Aifred-Health/pharma_research_model).

Data availability: Data requests will be considered by the study sponsor for non-commercial applications, on the condition that this publication is referenced in publications using the dataset. Requests for data access should be directed to david.benrimoh@mcgill.ca and will be responded to within 30 days. Data requested will be transferred in a de-identified format using patient ID numbers.

Acknowledgements:
We thank patients, physicians, and local staff for participating in this study.

Authorship statement:
DB, SI, KP, MR, HM, conceived the study and revised the paper. DB conducted and supervised the study and analyses, and wrote the first draft of the paper. KW, MR, GG, SI, KP, CP, AA, and SQ assistant with data collection and manuscript review. KW, MR, GG, KP, SJ, TF, YB, RF and CP assisted with analysis and paper revision. JM, RF, CA, KP, and DB produced the machine learning model used in the study and revised the paper. JK, SP, SG, EM, ME, SR, FHE, and HM served on the steering committee and revised the paper. SG and EM served as statisticians who oversaw sample size calculations carried out by JS and reviewed the statistical analysis plan, which was approved by DB, MR and HM. AG, MF, SP, MS, SK, GP, BD, KL, MR, served as site primary investigators and revised the paper. SB, WS served on the data safety monitoring board and revised the paper. RMS provided the STAR questionnaire and revised the paper. HM provided supervision for the study, the analysis, and revised the paper.

REFERENCES:

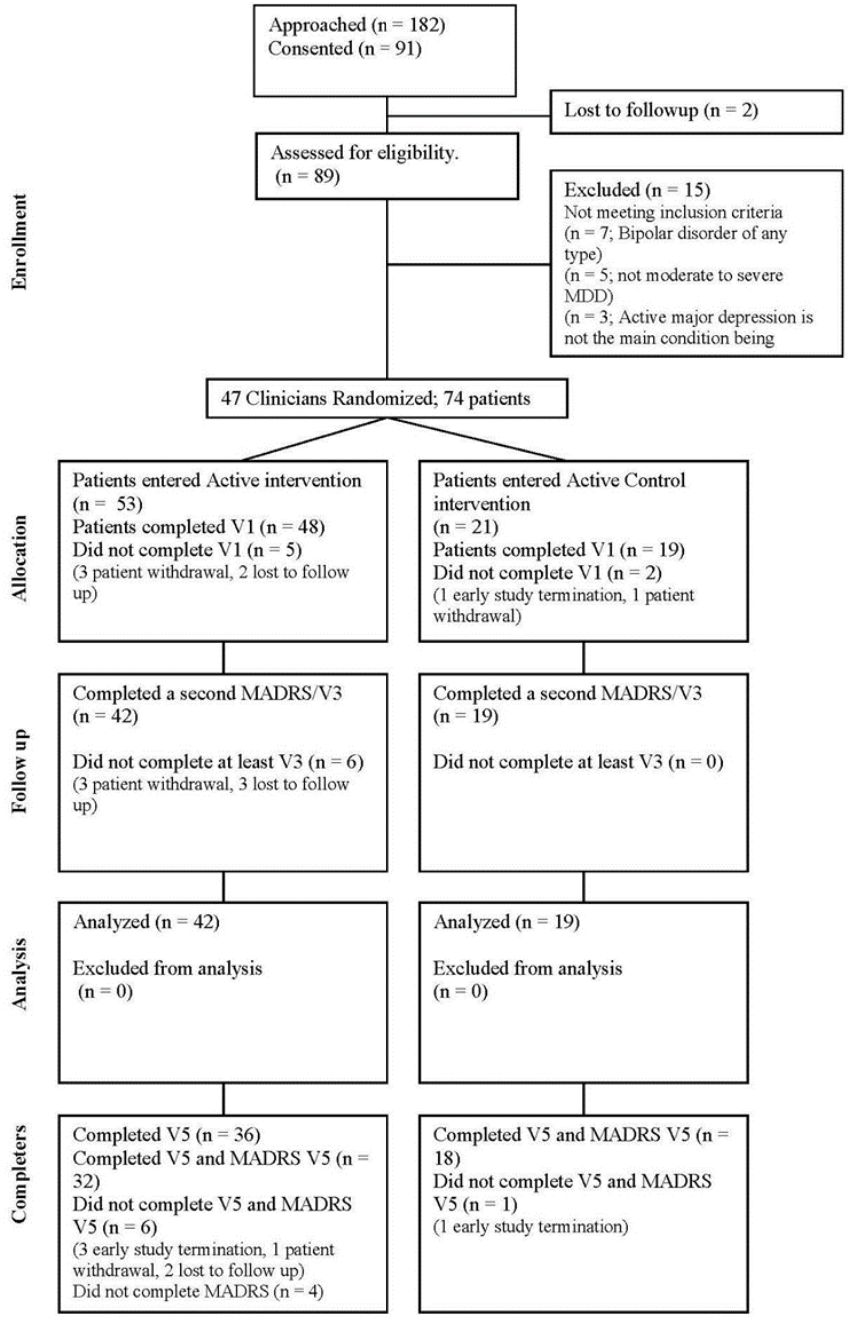
1. Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T. & Kessler, R. C. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J. Clin. Psychiatry* **76**, 155–162 (2015).
2. Health Organization, W. Depression and other common mental disorders: global health estimates. (2017).
3. Rush, A. J. *et al.* Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).
4. Kraus, C., Kadriu, B., Lanzenberger, R., Zarate, C. A., Jr & Kasper, S. Prognosis and improved outcomes in major depression: a review. *Transl. Psychiatry* **9**, 127 (2019).
5. Benrimoh, D. *et al.* Aifred Health, a Deep Learning Powered Clinical Decision Support System for Mental Health. in *The NIPS '17 Competition: Building Intelligent Systems* 251–287 (Springer International Publishing, 2018).
6. Mehltretter, J. *et al.* Analysis of Features Selected by a Deep Learning Model for Differential Treatment Selection in Depression. *Front Artif Intell* **2**, 31 (2019).
7. Squarcina, L., Villa, F. M., Nobile, M., Grisan, E. & Brambilla, P. Deep learning for the prediction of treatment response in depression. *J. Affect. Disord.* **281**, 618–622 (2021).
8. Poon, A. I. F. & Sung, J. J. Y. Opening the black box of AI-Medicine. *J. Gastroenterol. Hepatol.* **36**, 581–584 (2021).
9. Maslej, M. M., Kloiber, S., Ghassemi, M., Yu, J. & Hill, S. L. Out with AI, in with the psychiatrist: a preference for human-derived clinical decision support in depression care. *Transl. Psychiatry* **13**, 210 (2023).
10. Celi, L. A. *et al.* Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. *PLOS Digit Health* **1**, e0000022 (2022).
11. Schneider, F. *et al.* Insufficient depression treatment in outpatient settings. *Ger. Med. Sci.* **2**, Doc01 (2004).
12. Lisinski, A., Hieronymus, F., Eriksson, E. & Wallerstedt, S. M. Low SSRI dosing in clinical

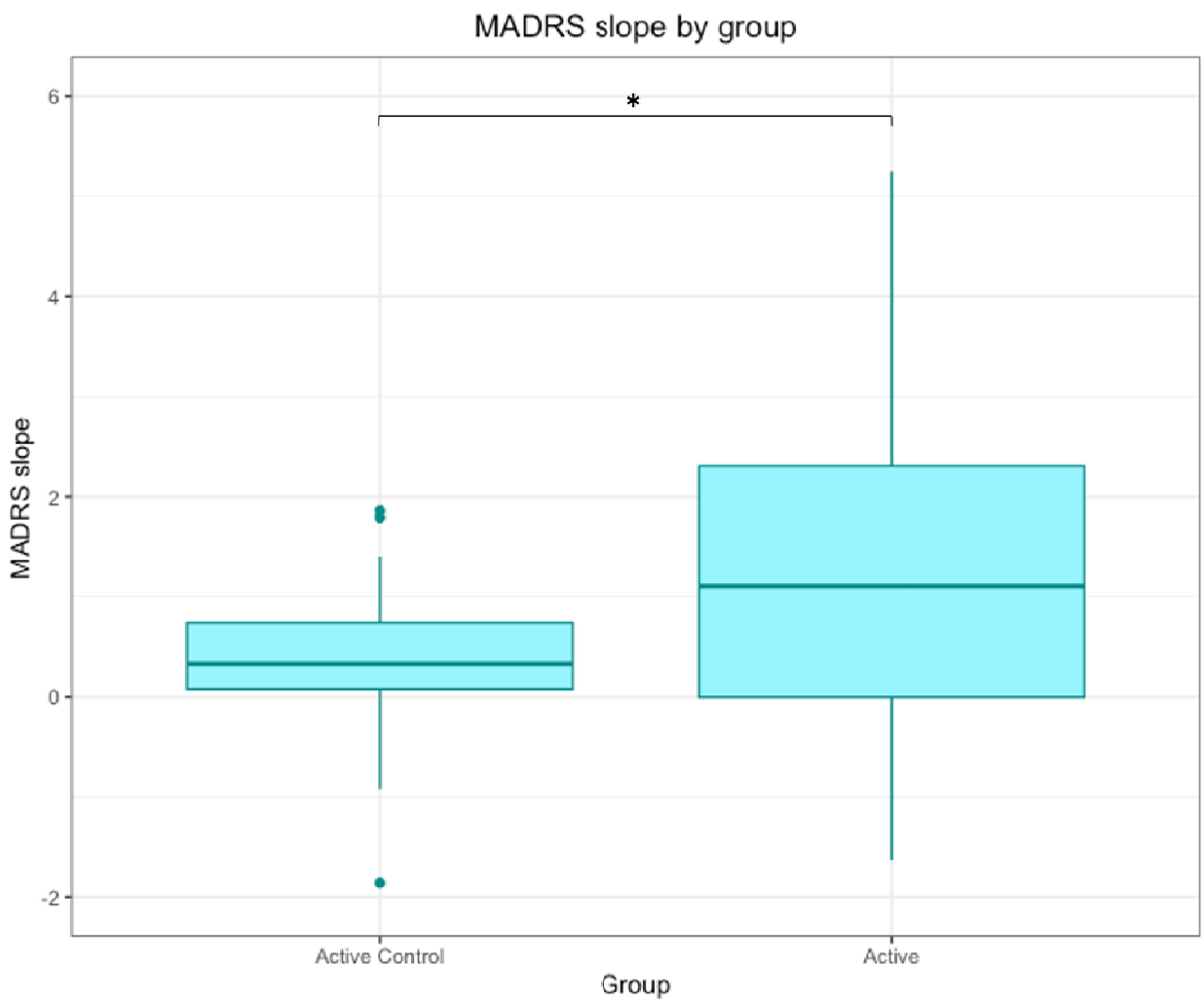
practice-a register-based longitudinal study. *Acta Psychiatr. Scand.* **143**, 434–443 (2021).

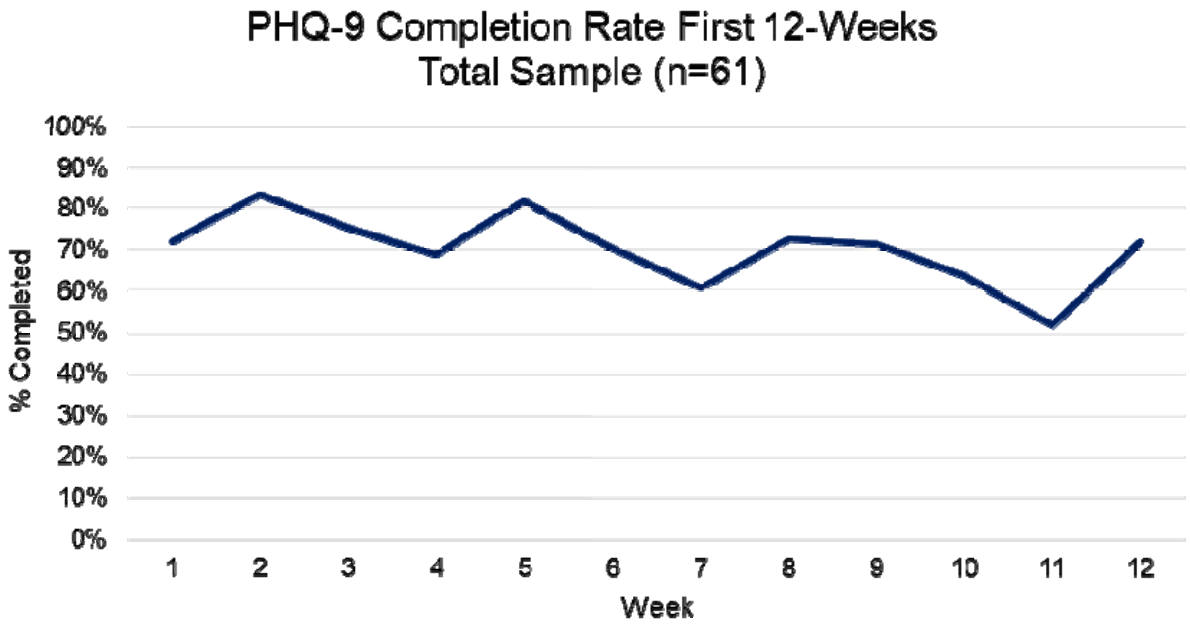
13. von Knorring, J. *et al.* Prospective study of antidepressant treatment of psychiatric patients with depressive disorders: treatment adequacy and outcomes. *BMC Psychiatry* **23**, 888 (2023).
14. Golden, G. *et al.* Applying artificial intelligence to clinical decision support in mental health: What have we learned? *Health Policy and Technology* 100844 (2024).
15. Kennedy, S. H. *et al.* Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 3. Pharmacological Treatments. *Can. J. Psychiatry* **61**, 540–560 (2016).
16. Benrimoh, D. *et al.* Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician--patient interaction. *BJPsych open* **7**, e22 (2021).
17. Tanguay-Sela, M. *et al.* Evaluating the perceived utility of an artificial intelligence-powered clinical decision support system for depression treatment using a simulation center. *Psychiatry Res.* **308**, 114336 (2022).
18. Popescu, C. *et al.* Evaluating the clinical feasibility of an artificial intelligence--powered, web-based clinical decision support system for the treatment of depression in adults: longitudinal feasibility study. *JMIR formative research* **5**, e31862 (2021).
19. Qassim, S. *et al.* A mixed-methods feasibility study of a novel AI-enabled, web-based, clinical decision support system for the treatment of major depression in adults. *Journal of Affective Disorders Reports* **14**, 100677 (2023).
20. Mehltretter, J. *et al.* Differential treatment Benet prediction for treatment selection in depression: A deep learning analysis of STAR*D and CO-MED data. *Comput. Psychiatr.* **4**, 61 (2020).
21. Kleinerman, A. *et al.* Treatment selection using prototyping in latent-space with application to depression treatment. *PLoS One* **16**, e0258400 (2021).

22. Benrimoh, D. *et al.* Towards Outcome-Driven Patient Subgroups: A Machine Learning Analysis Across Six Depression Treatment Studies. *Am. J. Geriatr. Psychiatry* **32**, 280–292 (2024).
23. Liu, X. *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ* **370**, m3164 (2020).
24. Cook, A. J., Delong, E., Murray, D. M., Vollmer, W. M. & Heagerty, P. J. Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core. *Clin. Trials* **13**, 504–512 (2016).
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. (American Psychiatric Publishing, 2013).
26. Sheehan, D. V. *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59 Suppl 20**, 22-33;quiz 34-57 (1998).
27. Asberg, M., Montgomery, S. A., Perris, C., Schalling, D. & Sedvall, G. A comprehensive psychopathological rating scale. *Acta Psychiatr. Scand. Suppl.* 5–27 (1978).
28. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
29. Perlman, K. *et al.* Development of a differential treatment selection model for depression on consolidated and transformed clinical trial datasets. *medRxiv* (2024) doi:10.1101/2024.02.19.24303015.
30. Benrimoh, D. *et al.* Development and Validation of a Deep-Learning Model for Differential Treatment Benefit Prediction for Adults with Major Depressive Disorder Deployed in the Artificial Intelligence in Depression Medication Enhancement (AIDME) Study. *arXiv [q-bio.NC]* (2024).
31. Fleischhacker, W. W. *et al.* Placebo or active control trials of antipsychotic drugs? *Arch. Gen. Psychiatry* **60**, 458–464 (2003).

32. Center for Drug Evaluation & Research. E10 choice of control group and related issues in clinical trials. *U.S. Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e10-choice-control-group-and-related-issues-clinical-trials (2020).
33. Byerly, M. J., Nakonezny, P. A. & Rush, A. J. The Brief Adherence Rating Scale (BARS) validated against electronic monitoring in assessing the antipsychotic medication adherence of outpatients with schizophrenia and schizoaffective disorder. *Schizophr. Res.* **100**, 60–69 (2008).
34. McIntyre, R. S. *et al.* Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale. *CMAJ* **173**, 1327–1334 (2005).
35. Kaneriya, S. H. *et al.* Predictors and Moderators of Remission With Aripiprazole Augmentation in Treatment-Resistant Late-Life Depression: An Analysis of the IRL-GRey Randomized Clinical Trial. *JAMA Psychiatry* **73**, 329–336 (2016).
36. Page, S. J. & Persch, A. C. Recruitment, retention, and blinding in clinical trials. *Am. J. Occup. Ther.* **67**, 154–161 (2013).
37. Hengartner, M. P. & Plöderl, M. Estimates of the minimal important difference to evaluate the clinical significance of antidepressants in the acute treatment of moderate-to-severe depression. *BMJ Evid Based Med* **27**, 69–73 (2022).
38. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
39. Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402–2410 (2016).
40. He, J. *et al.* The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* **25**, 30–36 (2019).


Approached (n = 182)
Consented (n = 91)
Lost to followup (n = 2)
Assessed for eligibility.
(n = 89)
Excluded (n = 15)
Not meeting inclusion criteria
(n = 7; Bipolar disorder of any type)
(n = 5; not moderate to severe MDD)
(n = 3; Active major depression is not the main condition being
Enrollment
47 Clinicians Randomized; 74 patients
Allocation
Patients entered Active intervention
(n = 53)
Patients completed V1 (n = 48)
Did not complete V1 (n = 5)
(3 patient withdrawal, 2 lost to follow up)
Patients entered Active Control intervention
(n = 21)
Patients completed V1 (n = 19)
Did not complete V1 (n = 2)
(1 early study termination, 1 patient withdrawal)
Follow up
Completed a second MADRS/V3
(n = 42)
Did not complete at least V3 (n = 6)
(3 patient withdrawal, 3 lost to follow up)
Completed a second MADRS/V3
(n = 19)
Did not complete at least V3 (n = 0)
Analysis
Analyzed (n = 42)
Excluded from analysis
(n = 0)
Analyzed (n = 19)
Excluded from analysis
(n = 0)
Completers
Completed V5 (n = 36)
Completed V5 and MADRS V5 (n = 32)
Did not complete V5 and MADRS V5 (n = 6)
(3 early study termination, 1 patient withdrawal, 2 lost to follow up)
Did not complete MADRS (n = 4)
Completed V5 and MADRS V5 (n = 18)
Did not complete V5 and MADRS V5 (n = 1)
(1 early study termination)


MADRS slope by group
*
6
4
2
0
-2
MADRS slope
Active Control
Active
Group


PHQ-9 Completion Rate First 12-Weeks
Total Sample (n=61)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% Completed
1
2
3
4
5
6
7
8
9
10
11
12
Week