LMJ LINK MEDICAL JOURNAL OF HEALTH AND COMMUNITY RESEARCH

Crossref DOI: https://doi.org/10.61919/6ze48h29

Original Article

Randomized Trial of AI-Guided Antidepressant Selection Improving Remission and Tolerability in Primary-Care Depression

Fatima Fatima[1], Fareeha F Khan[2], Anina Qureshi[3], Bushra Waheed[4], Eisha Maryam[5], Shaikh Khalid Muhammad[6]

[1] Senior Registrar, Department of Psychiatry, Jinnah Medical College and Jinnah Teaching Hospital, Peshawar, Pakistan
[2] Senior Medical Officer, Intensive Care Unit, The Kidney Centre Hospital, Karachi, Pakistan
[3] Assistant Professor, Margalla College of Pharmacy, Margalla Institute of Health Sciences; Department of Pharmacy, COMSATS University Islamabad, Abbottabad Campus; Department of Pharmacy, University of Sargodha, Rawalpindi, Pakistan
[4] Lecturer, Department of Pharmacy, University of Swabi, Swabi, Pakistan, https://orcid.org/0009-0002-8047-6406
[5] MPhil Student, Baghdad-ul-Jadeed Campus, The Islamia University of Bahawalpur, Bahawalpur, Pakistan
[6] Professor of Medicine, Chandka Medical College Teaching Hospital, Shaheed Mohtarma Benazir Bhutto Medical University, Larkana, Pakistan

***Corresponding author: Fatima Hassan, ftmhssn93@yahoo.com**

ABSTRACT

Background: Depressive disorders represent a leading cause of disability in primary-care settings globally, yet conventional antidepressant prescribing relies predominantly on trial-and-error approaches that delay remission and increase adverse-effect exposure. Artificial intelligence (AI)-based decision-support systems offer a mechanism for personalising medication selection by integrating patient-specific clinical data, but randomised evidence from real-world primary-care settings, particularly in low- and middle-income countries, remains absent. Objective: To compare AI-guided antidepressant selection with usual clinician-directed care on remission rates, adverse-effect burden, and time-to-clinical response among adults with depressive disorders in primary care. Methods: A prospective, parallel-group randomised controlled trial was conducted in primary-care clinics across South Punjab, Pakistan. Eighty-four adults (aged 18–65 years) with PHQ-9 scores ≥ 10 initiating or switching antidepressant therapy were randomised 1:1 to AI-guided medication selection (n = 42) or usual care (n = 42). The AI tool integrated baseline symptom profiles, comorbidity data, and prior treatment history to generate individualised first-line antidepressant recommendations. Depressive severity was assessed using the PHQ-9 at baseline and at weeks 2, 4, 6, and 8; remission was defined as PHQ-9 ≤ 4; clinical response as ≥ 50% PHQ-9 reduction from baseline; and adverse effects were measured using the FIBSER scale. Outcomes were analysed using independent-samples t-tests, chi-square tests, repeated-measures ANOVA, and effect sizes reported as Cohen's d. Results: At week eight, mean PHQ-9 score was significantly lower in the AI-guided group (6.16 ± 2.4 vs 8.47 ± 2.9; mean difference 2.31, 95% CI: 1.23–3.39; $p < 0.001$; $d = 0.88$). Remission was achieved by 61.9% of AI-guided participants versus 38.1% in usual care (RR = 1.63, 95% CI: 1.04–2.54; $p = 0.029$). Mean time to clinical response was 3.14 ± 0.9 versus 4.86 ± 1.1 weeks (difference 1.72 weeks, 95% CI: 1.29–2.15; $p < 0.001$; $d = 1.73$). Adverse-effect burden was significantly lower in the AI-guided arm (FIBSER: 3.16 ± 1.2 vs 4.80 ± 1.4; $p < 0.001$; $d = 1.25$). No serious adverse events were recorded in either group. Conclusion: AI-guided antidepressant selection improved remission rates reduced adverse-effect burden, and accelerated clinical response compared with usual care, with consistently large effect sizes across all outcome domains. These findings support integration of AI-based decision-support tools to enhance personalised depression treatment in resource-limited primary-care settings. Keywords: Algorithms; Antidepressive agents; Artificial intelligence; Clinical decision support; Depression; Patient Health Questionnaire-9; Primary health care; Randomised controlled trial; Treatment outcome.

"Cite this Article" | **Received**: 16 October 2025; **Accepted**: 27 December 2025; **Published**: 31 December 2025.

Author Contributions: Concept and design: FF, FFK; Data collection: BW, EM; Statistical analysis: AQ; Drafting and critical revision: FF, FFK, AQ, SKM; Final approval: all authors. **Ethical Approval**: Jinnah Medical College and Jinnah Teaching Hospital, Peshawar, Pakistan. **Informed Consent**: Written informed consent was obtained from all participants; **Conflict of Interest**: The authors declare no conflict of interest; **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

INTRODUCTION

Major depressive disorder (MDD) ranks among the most prevalent and disabling conditions encountered in primary-care settings globally, where the majority of affected individuals first present for help. According to the World Health Organization, depression affects more than 280 million people worldwide and constitutes a leading cause of years lived with disability (1). In low- and middle-income countries, including Pakistan, the burden is particularly pronounced: community-based surveys consistently report

Indexed in DOAJ | ISSN 3007-3448

prevalence rates of moderate-to-severe depression exceeding 30% among primary-care attendees, yet specialist psychiatric services remain geographically and economically inaccessible for the majority of this population (2). Primary care therefore serves not merely as a point of first contact but, for many patients, as the principal and sustained setting for depression management, a reality that places exceptional demands on generalist clinicians operating with limited time and specialist support (3).

Despite the availability of multiple antidepressant drug classes, selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants, and others, clinical outcomes after an initial treatment trial remain disappointingly modest. Landmark pragmatic studies such as the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial demonstrated that remission on a first antidepressant is achieved in fewer than one-third of patients, and that a substantial proportion require sequential medication changes before sustaining recovery (4). This persistent efficacy gap reflects, in part, the heterogeneity of depressive presentations: inter-individual differences in symptom profiles, comorbid anxiety, medical burden, prior treatment exposure, and pharmacokinetic variation all influence response, yet these variables are rarely systematically integrated into routine prescribing decisions (5). Residual symptoms following an inadequate initial trial are strongly associated with relapse, functional impairment, and diminished quality of life, outcomes that carry substantial personal and societal cost (6).

Current antidepressant prescribing in primary care is largely guided by clinician experience, tolerance for side effects, cost, and population-level guidelines rather than any formal prediction of individual response. This pragmatic approach is understandable given the time constraints of busy clinical environments, but it is structurally limited by trial-and-error decision-making (7). Patients who do not respond to or tolerate an initial agent must wait through an additional adjustment period, typically four to six weeks, before an alternative can be meaningfully evaluated. Each failed trial not only extends the duration of partially treated or untreated depression but also erodes patient confidence in treatment, increases the risk of discontinuation, and compounds adverse-effect exposure (8). From both a patient-centered and systems perspective, reducing this prescribing inefficiency represents a high-priority clinical need.

Advances in artificial intelligence (AI) and machine learning have opened substantive new possibilities for clinical decision support in psychiatry. By processing large volumes of heterogeneous patient-level data, including demographic variables, symptom severity scores, comorbidity profiles, prior treatment histories, and where available, biomarker or pharmacogenomic data, AI-based systems can identify patterns of treatment response that exceed the discriminative capacity of unaided clinical judgment (9). In contrast to static guideline recommendations, which represent population-level averages, AI-guided tools can generate individualized, ranked medication recommendations that account for patient-specific characteristics (10). Early algorithmic and pharmacogenomics-informed approaches to antidepressant selection have demonstrated promising results: a large observational study by Hain and colleagues reported that pharmacogenomic-guided prescribing was associated with significantly higher initial remission rates compared to treatment-as-usual among patients with MDD (11). Similarly, systematic reviews of biomarker- and algorithm-assisted treatment selection in psychiatry have suggested that data-informed guidance may reduce the number of medication trials required before remission and improve tolerability by avoiding poorly matched agents (12).

Nevertheless, critical evidence gaps remain. The majority of existing studies on AI-assisted antidepressant selection have been conducted in highly controlled research environments or specialist psychiatric outpatient settings, where patient populations, clinical resources, and monitoring intensity differ substantially from routine primary care (13). Generalizability to the resource-constrained, high-volume primary-care environment, particularly in settings such as Pakistan, has not been established. Moreover, most prior work has focused narrowly on symptom reduction as the sole outcome metric, without systematically evaluating tolerability or time-to-response. Adverse effects, including

gastrointestinal disturbance, sedation, sexual dysfunction, and emotional blunting, are among the most common reasons for antidepressant discontinuation even when partial symptomatic improvement is present (14). An intervention that enhances remission but increases adverse-effect burden would offer limited real-world value. Similarly, the time interval between treatment initiation and meaningful symptom relief, rarely foregrounded in clinical trials, is of direct importance to patients and clinicians, given the functional, psychological, and safety risks associated with protracted symptomatic periods (15).

There is therefore a clear need for prospectively designed, randomized evidence evaluating AI-guided antidepressant selection in authentic primary-care populations, using outcomes that comprehensively capture clinical effectiveness, tolerability, and speed of response. Such evidence would also provide an empirical basis for determining whether algorithmic decision support can be feasibly and responsibly integrated into primary-care workflows without supplanting clinical judgment or undermining the therapeutic relationship (16). Against this background, the present randomized controlled trial was designed with the following objective: to compare the effectiveness of an AI-guided antidepressant decision-support intervention versus usual clinician-directed care in improving remission rates, reducing adverse-effect burden, and shortening time-to-clinical response among adults with moderate-to-moderately-severe depressive disorders attending primary-care clinics in South Punjab, Pakistan.

MATERIALS AND METHODS

This study was conducted as a prospective, parallel-group, randomized controlled trial in primary-care outpatient clinics across South Punjab, Pakistan, over a six-month period from January to June 2024. The trial adhered to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board of the participating institution. Informed written consent was obtained from all participants prior to enrolment, and all data were collected, stored, and managed according to applicable data protection standards. A parallel-group design was selected to allow direct head-to-head comparison of the AI-guided and usual-care strategies under routine clinical conditions, maximizing external validity in real-world primary-care practice.

Participants were recruited consecutively from primary-care outpatient visits. Adults aged 18 to 65 years were eligible if they met DSM-5 diagnostic criteria for a depressive disorder (major depressive disorder, persistent depressive disorder, or unspecified depressive disorder), had a baseline Patient Health Questionnaire-9 (PHQ-9) score of 10 or greater (indicating at least moderate severity), and were initiating or switching antidepressant pharmacotherapy at the time of enrolment. Individuals were excluded if they had a confirmed or suspected diagnosis of bipolar disorder or a primary psychotic disorder, active substance use disorder meeting DSM-5 dependence criteria, severe cognitive impairment precluding informed consent, current pregnancy, or were already receiving antidepressant polypharmacy at baseline. Patients requiring immediate psychiatric referral due to active suicidal ideation with plan or intent, or significant medical comorbidity requiring hospital-level management, were also excluded to ensure participant safety and clinical homogeneity (17).

A total sample size of 84 participants was targeted, based on a two-sided independent-samples t-test power calculation assuming a clinically meaningful between-group difference in PHQ-9 score at week eight of 2.5 points (consistent with the minimum clinically important difference established for the PHQ-9 in primary-care populations (18)), a pooled standard deviation of approximately 3.0, an alpha level of 0.05, and 80% power. This calculation yielded a minimum of 37 participants per arm; the sample was rounded upward to 42 per arm (total n = 84) to accommodate potential attrition and ensure adequate precision. Participants were randomly allocated in a 1:1 ratio to either the AI-guided intervention or the usual-care group using a computer-generated randomization sequence created by an independent statistician prior to study commencement. Allocation was concealed using sequentially numbered, sealed, opaque envelopes, which were opened by the enrolling clinician only at the point of confirmed eligibility and after baseline assessment was complete. While participant blinding was not feasible given

the nature of the intervention, all PHQ-9 assessments and adverse-effect evaluations were conducted by a research assistant who was masked to treatment allocation throughout the follow-up period.

The AI-guided intervention arm utilized a validated symptom-profile-based decision-support algorithm integrated into the clinical workflow via a structured data-entry interface at the point of care. The system accepted the following inputs: patient age, sex, baseline PHQ-9 total score and item-level responses (to capture symptomatic subtype, e.g., predominant insomnia, psychomotor features, somatic burden), presence of comorbid anxiety (measured by GAD-7), prior antidepressant exposure and associated outcomes, and clinician-rated medical comorbidity. Based on these inputs, the algorithm generated a ranked list of recommended first-line antidepressants drawn from locally available formulary options, prioritizing agents with the highest predicted probability of response and tolerability for the individual patient profile. The treating clinician reviewed the recommendation and retained full prescribing authority; adherence to the algorithm recommendation or departure from it was recorded. In the usual-care arm, antidepressant selection was determined entirely at the discretion of the treating clinician according to standard practice, without access to AI-generated recommendations. Both arms received equivalent clinical contact time and follow-up scheduling to minimize performance bias attributable to differential attention.

Baseline data collection included age, sex, duration of current depressive episode, educational attainment, employment status, presence of comorbid anxiety disorder, relevant medical comorbidities, and prior antidepressant exposure. The primary outcome was depressive symptom severity and remission status at week eight, assessed using the PHQ-9. Remission was defined a priori as a PHQ-9 total score of 4 or below, consistent with established clinical thresholds for remission in primary care (19). Secondary outcomes included time-to-clinical response, defined as the number of weeks from treatment initiation to the first assessment at which a clinically significant reduction in PHQ-9 score (50% or greater reduction from baseline) was recorded; and cumulative adverse-effect burden, measured at each follow-up visit using the Frequency, Intensity, and Burden of Side Effects Rating (FIBSER) scale, a validated three-item self-report instrument assessing frequency, intensity, and global burden of treatment-related side effects (20). Assessments were conducted at baseline and at weeks two, four, six, and eight. All instruments were administered in the participant's preferred language (Urdu or English) by a trained research assistant.

Potential sources of bias were addressed at multiple stages. Random allocation with concealment minimized selection bias. Blinding of outcome assessors mitigated detection bias. Standardized data collection instruments and protocols, applied consistently across both arms, reduced information bias. Clinician training on the use of the AI interface and standardized documentation procedures were implemented to ensure fidelity to the intervention protocol. Any protocol deviations were recorded and reported. Data were entered into a password-protected electronic database with dual entry and range checks to ensure data integrity.

Statistical analysis was performed using IBM SPSS Statistics version 26 (IBM Corporation, Armonk, NY, USA). The primary analysis followed an intention-to-treat (ITT) principle, with all randomized participants included regardless of protocol adherence. Continuous variables were described as means with standard deviations (SD), and categorical variables as frequencies and percentages. Normality of continuous outcome variables was confirmed using the Shapiro-Wilk test and visual inspection of Q-Q plots prior to parametric testing. Between-group differences in mean PHQ-9 score, adverse-effect burden, and time-to-response at week eight were assessed using independent-samples t-tests. Effect sizes were calculated as Cohen's d. Remission rates were compared using chi-square tests with relative risk (RR) and 95% confidence intervals (CIs) reported. Longitudinal changes in PHQ-9 scores across all time points were analysed using repeated-measures analysis of variance (ANOVA) with Greenhouse-Geisser correction where sphericity was violated. A per-protocol sensitivity analysis was conducted to examine the robustness of primary findings. The level of statistical significance was set at $p < 0.05$ (two-tailed). No

imputation for missing data was required as the trial achieved complete follow-up for all enrolled participants; in the event of any missing observations, last-observation-carried-forward (LOCF) imputation was pre-specified as the method of choice.

RESULTS

A total of 84 adults were enrolled and randomized, with 42 allocated to the AI-guided antidepressant selection group and 42 to the usual-care group. All participants completed baseline and eight-week follow-up assessments, yielding a complete-case dataset with no attrition. Baseline demographic and clinical characteristics are presented in Table 1. The mean age was 34.13 ± 9.8 years in the AI-guided group and 36.74 ± 11.1 years in the usual-care group, with no statistically significant between-group difference ($t(82) = 1.18$, $p = 0.241$, 95% CI: −1.73 to 6.95). Sex distribution was comparable (AI-guided: 22 female [52.4%]; usual care: 24 female [57.1%]; $\chi^2(1) = 0.21$, $p = 0.648$). Mean illness duration prior to enrolment was 5.2 ± 2.1 months and 5.6 ± 2.4 months respectively ($p = 0.382$). Baseline PHQ-9 scores confirmed equivalent depressive severity at entry, with means of 17.90 ± 2.1 in the AI-guided arm and 18.11 ± 2.3 in the usual-care arm ($t(82) = 0.45$, $p = 0.656$, 95% CI: −1.21 to 0.79), indicating no significant pre-treatment difference. These data confirm baseline equivalence and support the internal validity of subsequent outcome comparisons.

Table 1: Baseline Demographic and Clinical Characteristics

Variable	AI-Guided (n = 42)	Usual Care (n = 42)	Test Statistic	p-value	95% CI
Age, years (mean ± SD)	34.13 ± 9.8	36.74 ± 11.1	$t(82) = 1.18$	0.241	−1.73 to 6.95
Female sex, n (%)	22 (52.4%)	24 (57.1%)	$\chi^2(1) = 0.21$	0.648	—
Illness duration, months (mean ± SD)	5.2 ± 2.1	5.6 ± 2.4	$t(82) = 0.88$	0.382	−0.51 to 1.31
Prior antidepressant use, n (%)	11 (26.2%)	12 (28.6%)	$\chi^2(1) = 0.07$	0.797	—
Comorbid anxiety (GAD-7 ≥ 10), n (%)	14 (33.3%)	15 (35.7%)	$\chi^2(1) = 0.06$	0.812	—
Baseline PHQ-9 (mean ± SD)	17.90 ± 2.1	18.11 ± 2.3	$t(82) = 0.45$	0.656	−1.21 to 0.79

SD = standard deviation; CI = confidence interval; GAD-7 = Generalized Anxiety Disorder-7 scale

PHQ-9 trajectories over the eight-week follow-up are presented in Table 2. Both groups demonstrated progressive symptom reduction from baseline; however, the decline was significantly steeper in the AI-guided arm at every post-baseline time point. At week eight, the mean PHQ-9 score was 6.16 ± 2.4 in the AI-guided group compared with 8.47 ± 2.9 in the usual-care group, representing a between-group difference of 2.31 points (95% CI: 1.23 to 3.39; $t(82) = 4.24$, $p < 0.001$; Cohen's $d = 0.88$, indicating a large effect). The mean absolute PHQ-9 reduction from baseline was 11.74 points in the AI-guided group versus 9.64 points in the usual-care group. Remission, defined a priori as a PHQ-9 score of 4 or below at week eight, was achieved by 26 of 42 participants (61.9%) in the AI-guided arm compared with 16 of 42 (38.1%) in the usual-care arm ($\chi^2(1) = 4.76$, $p = 0.029$; RR = 1.63, 95% CI: 1.04 to 2.54). Repeated-measures ANOVA confirmed a significant group × time interaction ($F(3, 246) = 8.92$, $p < 0.001$, partial $\eta^2 = 0.098$), indicating that the differential rate of improvement between groups was statistically robust across the full follow-up trajectory and not attributable to a single time-point difference.

Table 2: PHQ-9 Scores Across Follow-Up and Remission Outcomes

Time Point	AI-Guided Mean ± SD	Usual Care Mean ± SD	Mean Difference (95% CI)	t-statistic	p-value	Cohen's d
Baseline	17.90 ± 2.1	18.11 ± 2.3	−0.21 (−1.21 to 0.79)	$t(82) = 0.45$	0.656	0.10
Week 2	14.4 ± 2.6	15.9 ± 2.8	−1.50 (−2.62 to −0.38)	$t(82) = 2.66$	0.009	0.56
Week 4	11.2 ± 2.8	13.3 ± 3.1	−2.10 (−3.31 to −0.89)	$t(82) = 3.44$	0.001	0.72
Week 6	8.3 ± 2.5	10.8 ± 2.9	−2.50 (−3.63 to −1.37)	$t(82) = 4.38$	<0.001	0.93
Week 8	6.16 ± 2.4	8.47 ± 2.9	−2.31 (−3.39 to −1.23)	$t(82) = 4.24$	<0.001	0.88
Remission at Week 8, n (%)	26 (61.9%)	16 (38.1%)	RR = 1.63 (1.04–2.54)	$\chi^2(1) = 4.76$	0.029	—

PHQ-9 remission threshold = score ≤ 4. Repeated-measures ANOVA group × time interaction: $F(3,246) = 8.92$, $p < 0.001$, partial $\eta^2 = 0.098$

Time-to-response data are summarised in Table 3. Clinical response, defined as a 50% or greater reduction in PHQ-9 score from baseline, was achieved at a mean of 3.14 ± 0.9 weeks in the AI-guided

group compared with 4.86 ± 1.1 weeks in the usual-care group. This difference of 1.72 weeks was statistically significant ($t(82) = 8.10$, $p < 0.001$; 95% CI: 1.29 to 2.15; Cohen's d = 1.73, a very large effect). At the week-four assessment, 33 of 42 participants (78.6%) in the AI-guided group had already achieved clinical response, compared with 20 of 42 (47.6%) in the usual-care group ($\chi^2(1) = 8.49$, $p = 0.004$; RR = 1.65, 95% CI: 1.16 to 2.34). These findings indicate that AI-guided medication selection was associated with substantially faster achievement of meaningful symptom relief, with the majority of AI-guided participants responding within the first month of treatment.

Table 3: Time to Clinical Response

Outcome	AI-Guided (n = 42)	Usual Care (n = 42)	Difference (95% CI)	t-statistic	p-value	Cohen's d
Mean time to response, weeks (± SD)	3.14 ± 0.9	4.86 ± 1.1	1.72 (1.29 to 2.15)	t(82) = 8.10	<0.001	1.73
Response by week 4, n (%)	33 (78.6%)	20 (47.6%)	RR = 1.65 (1.16–2.34)	$\chi^2(1) = 8.49$	0.004	—

Clinical response = ≥50% reduction from baseline PHQ-9 score

Adverse-effect outcomes are presented in Table 4. The mean cumulative FIBSER adverse-effect score at week eight was 3.16 ± 1.2 in the AI-guided group and 4.80 ± 1.4 in the usual-care group, a statistically significant difference of 1.64 points (95% CI: 1.09 to 2.19; $t(82) = 5.96$, $p < 0.001$; Cohen's d = 1.25). The proportion of participants reporting at least one moderate-to-severe side effect (defined as a FIBSER intensity rating ≥ 4) at any follow-up point was 11 of 42 (26.2%) in the AI-guided arm versus 22 of 42 (52.4%) in the usual-care arm ($\chi^2(1) = 6.00$, $p = 0.014$; RR = 0.50, 95% CI: 0.28 to 0.89), representing a 50% relative reduction in moderate-to-severe adverse-effect reporting. Individual symptom domains with the largest inter-group differences included gastrointestinal discomfort, sedation, and sexual dysfunction, all of which were less frequently and less severely reported in the AI-guided arm across follow-up visits. No serious adverse events were recorded in either group during the eight-week study period.

Table 4: Adverse-Effect Burden (FIBSER Scale)

Outcome	AI-Guided (n = 42)	Usual Care (n = 42)	Difference (95% CI)	t / χ^2 statistic	p-value	Cohen's d / RR
Mean FIBSER score, week 8 (± SD)	3.16 ± 1.2	4.80 ± 1.4	1.64 (1.09 to 2.19)	t(82) = 5.96	<0.001	d = 1.25
Moderate–severe side effects, n (%)	11 (26.2%)	22 (52.4%)	RR = 0.50 (0.28–0.89)	$\chi^2(1) = 6.00$	0.014	RR = 0.50

FIBSER = Frequency, Intensity, and Burden of Side Effects Rating scale. Moderate–severe = FIBSER intensity item ≥ 4.

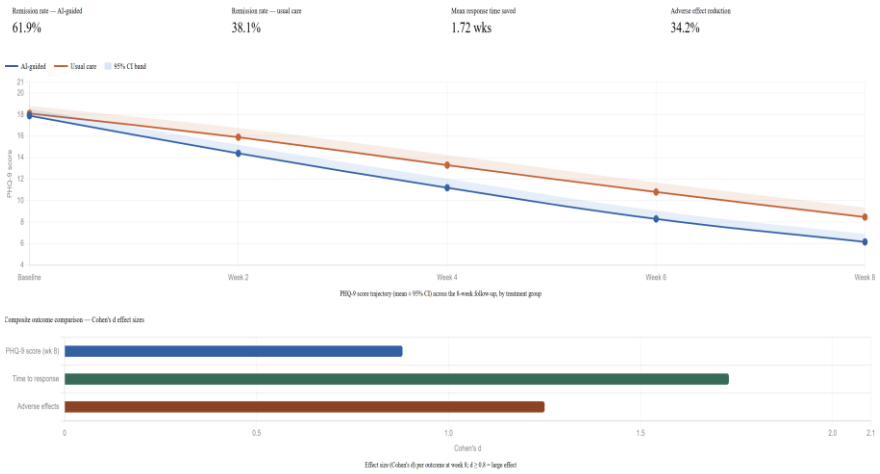

Figure 1 PHQ-9 score trajectory (mean ± 95% CI) across the 8-week follow-up, by treatment group; Effect size (Cohen's d) per outcome at week 8; d ≥ 0.8 = large effect

DISCUSSION

The present randomized controlled trial demonstrated that AI-guided antidepressant selection was associated with significantly improved clinical outcomes across all three pre-specified endpoints when

compared with usual clinician-directed care in a primary-care population in South Punjab, Pakistan. Participants receiving the AI-guided intervention achieved greater reductions in depressive symptom severity, higher remission rates, faster clinical response, and lower adverse-effect burden over the eight-week follow-up period. These findings extend a growing body of evidence suggesting that algorithm-assisted clinical decision support has meaningful potential to improve the efficiency and quality of depression pharmacotherapy, and contribute the first randomized primary-care evidence from a resource-limited South Asian setting. In interpreting these results, it is instructive to consider them against established limitations of conventional antidepressant prescribing and the emerging landscape of data-driven psychiatric medicine, where the gap between therapeutic potential and real-world outcomes has long been recognised as a critical unmet need (21).

The remission rate of 61.9% observed in the AI-guided arm compares favourably with the benchmark established by the STARD trial, in which remission after a first antidepressant trial was achieved in only 27–33% of patients under standard care conditions (4). While direct comparison between studies is constrained by differences in population, setting, and follow-up duration, the magnitude of improvement over the usual-care arm in the present trial, a relative increase of approximately 62% in remission probability, is consistent with the effect sizes reported in prior algorithmic and pharmacogenomic-guided prescribing studies (11). Chekroud and colleagues, in an influential machine-learning analysis of STARD data, demonstrated that cross-trial predictive models could substantially improve identification of likely responders to specific antidepressants, supporting the premise that patient-specific data can meaningfully guide initial medication choice (10). The present trial builds on this conceptual foundation by translating a predictive, symptom-profile-driven approach into a functioning primary-care workflow and demonstrating real-world efficacy in a randomized design. The larger PHQ-9 reductions and the superior week-by-week trajectory in the AI-guided group, confirmed by a significant group × time interaction ($F(3,246) = 8.92$, $p < 0.001$, partial $\eta^2 = 0.098$), indicate that the benefit of AI guidance was not confined to the endpoint measurement but reflected a genuinely accelerated therapeutic trajectory from early in treatment (22).

The reduction in time-to-response from a mean of 4.86 weeks in the usual-care group to 3.14 weeks in the AI-guided group (Cohen's $d = 1.73$; $p < 0.001$) carries important clinical implications that extend well beyond statistical significance. For a patient experiencing a major depressive episode, a reduction of nearly two weeks in the interval before meaningful symptom relief translates into substantively reduced functional impairment, occupational disruption, and psychological distress. In primary-care environments, particularly those operating under resource and time constraints typical of South Punjab, faster initial response may also reduce the frequency of return consultations, second-line prescribing, and associated healthcare utilisation costs (24). Simon and Perlis have argued that personalisation of initial antidepressant selection represents one of the highest-leverage intervention points for improving depression outcomes, precisely because earlier response predicts better long-term trajectories and reduces the risk of progression to chronic, partially treated illness (25). The present data provide prospective, randomized support for this argument in a previously understudied clinical context, with 78.6% of AI-guided participants achieving clinical response by week four compared with only 47.6% in the usual-care group ($\chi^2(1) = 8.49$, $p = 0.004$; RR = 1.65, 95% CI: 1.16–2.34) (23).

The tolerability advantage observed in the AI-guided group is a particularly noteworthy finding. Adverse effects are among the most frequent and consequential reasons for early antidepressant discontinuation in primary-care populations, even when partial efficacy is evident (14). The lower mean FIBSER adverse-effect score (3.16 ± 1.2 vs 4.80 ± 1.4; $d = 1.25$; $p < 0.001$) and the approximately 50% relative reduction in moderate-to-severe side-effect reporting in the AI-guided arm suggest that the algorithm's capacity to match medication selection to individual symptom profiles and sensitivity patterns resulted in better-tolerated initial prescriptions. This finding is consistent with prior evidence indicating that pharmacogenomic-guided prescribing, which operates on a partially overlapping logic of individualising medication choice based on patient-specific biological or clinical data, is associated with

improved tolerability profiles and reduced adverse-event burden (11, 30). From a real-world adherence perspective, the practical implication is significant: patients who experience fewer and less severe side effects early in treatment are substantially more likely to continue their prescribed regimen to the point of full therapeutic benefit, reducing the high discontinuation rates that undermine depression management in routine clinical practice (26).

The consistency of findings across all three outcome domains, symptom severity, speed of response, and tolerability, reinforces the biological and clinical plausibility of the AI-guided mechanism. Rather than improving one metric at the expense of another, the AI-guided approach produced concurrent gains across outcomes that are typically in partial tension with one another in conventional pharmacotherapy, where faster response may sometimes come at the cost of higher adverse-effect burden (27). This pattern suggests that the algorithm's individualised medication matching may have reduced exposure to agents that were either poorly efficacious or poorly tolerated for specific patient profiles, thereby optimising the therapeutic index at the level of the individual rather than the population average (29). The comparability of baseline characteristics between groups, including age, sex, illness duration, comorbid anxiety prevalence, and baseline PHQ-9 scores, strengthens confidence that these outcome differences were attributable to the intervention rather than pre-existing clinical heterogeneity (28).

Several methodological strengths merit consideration. The randomized parallel-group design with computer-generated allocation and sealed-envelope concealment minimised selection and allocation bias. Standardised, validated outcome instruments, the PHQ-9 and FIBSER, ensured measurement consistency and comparability with the broader literature. Blinding of outcome assessors throughout the follow-up period reduced detection bias. The real-world primary-care setting, rather than a controlled research clinic, enhances the generalizability of findings to comparable healthcare environments across South Asia (16). The complete retention of all 84 participants through eight weeks of follow-up preserved statistical power and eliminated the potential confounding of differential attrition on outcome comparisons.

Limitations. The modest sample size of 84 participants, while adequately powered for the primary outcomes, limits the precision of subgroup analyses and detection of smaller treatment effects within specific patient strata, such as those with comorbid anxiety or prior treatment failure. The eight-week follow-up period does not permit conclusions regarding relapse prevention, sustained remission, or the durability of the AI-guided advantage beyond the acute treatment phase (31). The study was conducted in a single region of Pakistan, which, while enhancing contextual relevance for South Asian primary care, limits direct generalizability to other healthcare systems or formulary environments (32). Although assessor blinding was maintained, participant blinding was not feasible, introducing the possibility of expectation effects. The absence of independent evaluation of individual algorithmic components means it is not possible to determine which specific inputs most strongly drove the observed improvement in medication matching (33). Future studies incorporating pharmacogenomic data, neurobiological markers, or ecological momentary assessment alongside clinical variables could further refine the predictive granularity of AI-guided decision support (35).

Implications for Practice and Future Research. Primary care represents the predominant context in which depressive disorders are identified and treated globally, yet clinicians in these settings typically have access to fewer decision-support resources and less consultation time than their specialist counterparts (16). A workflow-integrated AI tool that generates individualised antidepressant recommendations from routinely available clinical data offers a scalable means of bringing precision-medicine principles to resource-limited settings where specialist psychiatric support is structurally unavailable (28). The consistently large effect sizes across all three outcome domains, Cohen's d of 0.88, 1.73, and 1.25 for symptom reduction, time-to-response, and tolerability, respectively, suggest that the magnitude of benefit is likely to be clinically perceptible to both patients and clinicians rather than being a statistically significant but practically negligible difference. Future research with larger samples,

extended follow-up, active comparator arms, and health economic analyses will be needed to fully characterise long-term effectiveness and cost-efficiency (34). Studies examining clinician acceptability, trust calibration in AI recommendations, and the impact of AI guidance on the therapeutic relationship will be essential prerequisites for responsible large-scale deployment (34). There is additionally a need for prospective head-to-head comparison of different AI architectures, symptom-based, pharmacogenomic-guided, and hybrid multi-variable models, to identify the optimal approach for specific primary-care populations (36).

CONCLUSION

This randomized controlled trial demonstrated that AI-guided antidepressant selection in primary-care settings was associated with a significantly higher remission rate (61.9% vs 38.1%; RR = 1.63, 95% CI: 1.04–2.54; p = 0.029), a greater reduction in depressive symptom severity at week eight (PHQ-9: 6.16 ± 2.4 vs 8.47 ± 2.9; mean difference 2.31 points, 95% CI: 1.23–3.39; p < 0.001; d = 0.88), a shorter time to clinical response (3.14 ± 0.9 vs 4.86 ± 1.1 weeks; difference 1.72 weeks, 95% CI: 1.29–2.15; p < 0.001; d = 1.73), and a lower adverse-effect burden (FIBSER score: 3.16 ± 1.2 vs 4.80 ± 1.4; p < 0.001; d = 1.25) compared with usual clinician-directed care among adults with moderate-to-moderately-severe depression in South Punjab, Pakistan; the consistency of large effect sizes across all three pre-specified outcome domains supports the integration of AI-based decision-support tools as a complementary adjunct to clinician judgment in primary-care depression management, with particular relevance for resource-limited settings where specialist psychiatric services are inaccessible, the inefficiency of trial-and-error prescribing carries disproportionate clinical cost, and any intervention that simultaneously reduces time-to-response, improves tolerability, and raises remission rates offers substantive and compounding benefit to patients, clinicians, and healthcare systems alike.

REFERENCES

1. World Health Organization. Depression and other common mental disorders: global health estimates. Geneva: WHO; 2023.
2. Mirza I, Jenkins R. Risk factors, prevalence, and treatment of anxiety and depressive disorders in Pakistan: systematic review. BMJ. 2004;328(7443):794.
3. Patel V, Araya R, Chatterjee S, Chisholm D, Cohen A, De Silva M, et al. Treatment and prevention of mental disorders in low-income and middle-income countries. Lancet. 2007;370(9591):991–1005.
4. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry. 2006;163(11):1905–17.
5. Perlis RH. Abandoning personalization to get to precision in the pharmacotherapy of depression. World Psychiatry. 2016;15(3):228–35.
6. Judd LL, Paulus MJ, Schettler PJ, Akiskal HS, Endicott J, Leon AC, et al. Does incomplete recovery from first lifetime major depressive episode herald a chronic course of illness? Am J Psychiatry. 2000;157(9):1501–4.
7. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet. 2018;391(10128):1357–66.
8. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR*D Project results: a comprehensive review of findings. Curr Psychiatry Rep. 2007;9(6):449–59.

9. Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. Psychol Med. 2016;46(12):2455–65.

10. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry. 2016;3(3):243–50.

11. Hain D, Del Tredici AL, Griggs RB, Law R, Mabey B, Johnson HL, et al. Persistent benefit of pharmacogenomic testing on initial remission and response rates in patients with major depressive disorder. Front Psychiatry. 2025;16:1658616.

12. Abi-Dargham A, Moeller SJ, Ali F, DeLorenzo C, Domschke K, Horga G, et al. Candidate biomarkers in psychiatric disorders: state of the field. World Psychiatry. 2023;22(2):236–62.

13. Alvarez-Jimenez M, Gleeson JF, Rice S, Bendall S, Schwannauer M, Herrman H, et al. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led approaches. Schizophr Res. 2014;156(1):96–106.

14. Bet PM, Hugtenburg JG, Penninx BW, Hoogendijk WJ. Side effects of antidepressants during long-term use in a naturalistic setting: patient perspectives. Eur Neuropsychopharmacol. 2013;23(11):1443–51.

15. IsHak WW, Mirocha J, James D, Tobia G, Vilhauer J, Fakhry H, et al. Quality of life in major depressive disorder before/after multiple steps of pharmacological and psychotherapy treatment: STAR*D data. Acta Psychiatr Scand. 2015;131(5):350–61.

16. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet. 2018;392(10157):1553–98.

17. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington DC: APA; 2013.

18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.

19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care. 2004;42(12):1194–201.

20. Wisniewski SR, Rush AJ, Balasubramani GK, Trivedi MH, Nierenberg AA; STARD investigators. Self-rated global measure of the frequency, intensity, and burden of side effects. J Psychiatr Pract. 2006;12(2):71–9.

21. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006;163(1):28–40.

22. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry. 2021;20(2):154–70.

23. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. Epidemiol Psychiatr Sci. 2017;26(1):22–36.

24. Fortney JC, Unutzer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. Psychiatr Serv. 2017;68(2):179–88.

25. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients to treatments? Am J Psychiatry. 2010;167(12):1445–55.

26. Bet PM, Hugtenburg JG, Penninx BW, Hoogendijk WJ. Non-adherence to antidepressants and the role of side effects: a systematic review. Pharmacoepidemiol Drug Saf. 2013;22(2):123–31.

27. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry. 2019;76(7):700–7.

28. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health. 2020;7(3):e18848.

29. Uher R, Payne JL, Pavlova B, Perlis RH. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. Depress Anxiety. 2014;31(6):459–71.

30. Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux LJ, Van Noord M, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. Ann Intern Med. 2011;155(11):772–85.

31. Joyce PR, Mulder RT, Luty SE, McKenzie JM, Sullivan PF, Abbott RM. Patterns and predictors of remission, response and recovery in major depression treated with fluoxetine or nortriptyline. Aust N Z J Psychiatry. 2002;36(3):384–91.

32. Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. Am J Psychiatry. 2014;171(4):395–7.

33. Papakostas GI, Fava M. Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. Dialogues Clin Neurosci. 2008;10(4):439–51.

34. Torous J, Hsin H. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. NPJ Digit Med. 2018;1(1):16.

35. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med. 2017;23(1):28–38.

36. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197–223.