Weighting Schemes in Multi-Criteria Clinical Decision Support: A User Study of Physician Antidepressant Choice

Submitted to Multi-Objective Decision Making Workshop 2026

Abstract

Effective Multi-Criteria Decision Making (MCDM) in healthcare requires balancing algorithmic predictions with clinical agency. This challenge is particularly acute in Major Depressive Disorder (MDD) treatment, where clinicians must navigate complex trade-offs between remission probabilities and diverse, potentially competing, side-effect risks. While AI-driven Clinical Decision Support Systems (AI-CDSS) can assist in the prediction of individual outcomes, the optimal mechanism for aggregating these multi-dimensional criteria remains an open research challenge. In this work, we investigate how the locus of control in the aggregation mechanism affects clinical utility and decisions by evaluating three distinct weighting schemes: (1) an *Implicit Weighting* baseline (presenting raw probabilities), (2) a *Static Expert-Derived Weighting* (linear aggregation via fixed expert-based weights), and (3) a *Dynamic Clinician-Determined Weighting* (linear aggregation via adjustable weights). We implemented these schemes within a prototype decision support system for antidepressant selection and evaluated them through a user study with 22 physicians. Results demonstrate that the dynamic clinician-determined weighting scheme significantly enhances perceived clinical utility compared to alternatives ($p < 0.01$). Crucially, this approach also prompted the most frequent data-informed revision of the initial *unassisted* antidepressant choice (33.3% of cases), effectively catalyzing a shift toward a more informed treatment strategy among both specialists (psychiatrists) and primary care physicians. These findings suggest that AI-CDSS with MCDM should prioritize flexible aggregation mechanisms that empower clinical agency.

1 Introduction

Clinical Decision Support Systems (CDSS) have evolved from rule-based engines to sophisticated platforms leveraging Artificial Intelligence (AI) to empower data-driven clinical decisions [Aruldoss *et al.*, 2013; Zionts, 1979; Elhaddad and Hamam, 2024]. This evolution is particularly critical for mitigating the burden associated with Major Depressive Disorder (MDD), a condition characterized by high prevalence and significant morbidity [Benrimoh *et al.*, 2021; Triantaphyllou, 2000]. Despite the availability of pharmacological interventions, the current standard of care often relies on an "educated trial-and-error" approach [Dunn, 1964], where only a third of patients achieve remission after their first treatment [Trivedi *et al.*, 2004]. In this context, AI models have demonstrated significant potential in predicting individual treatment outcomes, including probabilities of remission [Thokala *et al.*, 2016; Charness *et al.*, 2012; Triantaphyllou, 2000] and specific adverse events such as weight gain, sexual dysfunction, and fatigue [Burra *et al.*, 2007; Ashton *et al.*, 2005].

Nevertheless, despite improvements in predictive accuracy, the translation of these models into clinical practice remains limited [Sarantopoulos *et al.*, 2024; Van Booven *et al.*, 2025]. A fundamental barrier to adoption is that clinical decision-making is rarely a single-objective optimization task. Instead, it is a Multi-Criteria Decision Making (MCDM) challenge involving potentially conflicting objectives [Abu-Mahfouz *et al.*, 2026; Popescu *et al.*, 2021; Bet *et al.*, 2013]. In antidepressant selection, this challenge is particularly acute as clinicians must delicately balance efficacy against the risk of side effects, trade-offs that vary significantly depending on the clinical context and patient preferences. While AI can accurately predict the *risk* associated with these outcomes, it cannot inherently determine their *subjective value* within the clinician-patient shared decision-making process [Elwyn *et al.*, 2012; Serretti *et al.*, 2010; Zhou *et al.*, 2022].

Inevitably, a core challenge in designing AI-CDSS is not merely prediction, but the aggregation of multi-dimensional outputs into actionable recommendations. Existing MCDM approaches in AI-CDSS typically fall into two extremes: (i) *Implicit Weighting*, also known as "Probabilities Alone (PA)", where systems present raw probabilities derived from computational models. While transparent, this approach forces clinicians to perform the challenging cognitive integration of these probabilities manually; and (ii) *Static Expert-Derived Weighting*, where systems aggregate model outputs using fixed rules, commonly a weighted sum model using expert-

defined linear weights [Najafi *et al.*, 2023; Aljohani, 2025]. While this approach is designed to reduce cognitive load, it effectively limits the clinician's ability to adapt the decision logic to patient-specific nuances. This creates an inherent tension between maintaining clinician agency at the cost of high cognitive burden (the PA scheme) or accepting streamlined processing at the cost of rigid, generalized decision logic (the static scheme).

To this end, in this study, we address this tension by proposing and evaluating a middle-ground solution: *Dynamic Clinician-Determined Weighting*. By decoupling risk prediction (AI-based) from subjective weighting alignment (Human-based), this approach allows clinicians to explicitly parameterize the trade-offs relevant to the current case and dynamically adjust them if needed [Martin *et al.*, 2022]. We compare this clinician-determined weighting scheme against both implicit and static alternatives through a controlled user study with 22 physicians. Participants interacted with an AI-CDSS prototype across diverse clinical scenarios, allowing us to assess how the locus of control in weighting mechanisms influences the perceived utility and data-informed clinical decisions.

2 Methodology

To study the impact of weighting schemes on clinical decision-making, we conducted a within-subjects clinician user study ($N = 22$) comparing three distinct weighting schemes implemented to support antidepressant selection. The study was designed to assess how the locus of control (implicit, expert-determined, or clinician-determined) influences clinicians' perceived utility and treatment selection.

2.1 Experimental Design and Procedure

The experiment employed a randomized, within-subjects design [Charness *et al.*, 2012]. Following informed consent and a standard demographic questionnaire, participants completed three distinct decision-making phases. In each phase, participants were presented with: (1) a vignette describing a patient with Major Depressive Disorder (MDD), and (2) one of the three examined weighting schemes integrated within an AI-CDSS. The pairing of clinical scenarios and weighting schemes was randomized to control for order and case-difficulty effects. Figure 1 presents a schematic view of the methodological process utilized in this study.

At each phase, participants first reviewed the patient profile. Crucially, *before* interacting with any form of decision support for that patient, they were asked to record their *unassisted* antidepressant choice based on the clinical vignette alone. They then interacted with a prototype AI-CDSS integrated with one of the three weighting schemes to generate recommendations and select a final antidepressant. Following each phase, participants completed a very short questionnaire containing only two questions: "How would you rate the clinical usefulness of the decision support you just experienced? (five point Likert-sclae: 1=not very useful, 5=very useful)" and "Did you change your initial treatment selection following the decision support? (Yes/No)". A concluding (exit) questionnaire was also administered to probe for comparative preferences and open-form qualitative feedback.

2.2 AI-CDSS Framework and Weighting Schemes

We developed a web-based AI-CDSS framework aimed at supporting MCDM among five common antidepressants (Escitalopram, Sertraline, Bupropion, Venlafaxine, and Mirtazapine). This set of antidepressants represents both pharmacological diversity and is the most frequently prescribed antidepressants worldwide [Lam, 2018]. To isolate the impact of the weighting mechanism, the framework was designed to simulate the output of a predictive AI model tailored for MDD treatment. The system presented realistic, pre-defined probability estimates for remission and the risk of three major side effects: weight gain, sexual dysfunction, and fatigue. These estimates, informed by pharmacological literature and expert consensus for three realistic patient profiles (see Appendix), were constructed to reflect plausible clinical profiles and trade-offs. The specific side effects were selected based on their high prevalence as drivers of non-adherence [Hunot *et al.*, 2007; Archibald, 2016; Popescu *et al.*, 2021].

We examined three distinct weighting schemes for aggregating these predictions:

Condition 1: Implicit Weighting (Probabilities Alone)

In this setting, the system displayed raw probability estimates for remission and side effects via interactive bar charts (Figure 2) and tabular data. No aggregated score was provided. This baseline mimics standard "dashboard"-like analytics, forcing the clinician to perform the MCDM aggregation implicitly (cognitively) without algorithmic assistance.

Condition 2: Static Expert-Derived Weighting

The Expert-Derived Weighting (EDW) scheme implements a weighted sum model to calculate an aggregated score S_j for each drug j, calculated as follows:

$$S_j = \frac{\sum_{i=1}^{n} w_i \cdot v_{ij}}{\sum_{i=1}^{n} w_i} \tag{1}$$

where v_{ij} is the normalized predicted value of outcome i for drug j, and w_i represents the weight assigned to that outcome. The five drugs were presented as a ranked list sorted by S_j, with explainability supported via stacked bar charts decomposing the contribution of each criterion to the total score (Figure 3). Importantly, the weights w_i were pre-defined by a panel of three senior psychiatrists who reviewed the patient profiles and deliberated to reach a consensus on the optimal weights for each case. The integration of static weights simulates the "black-box" nature of many AI-CDSSs, where value alignment is determined by external domain knowledge rather than the user.

Condition 3: Dynamic Clinician-Determined Weighting

The Clinician-Determined Weighting (CDW) scheme follows the same technical and visualization approach as the EDW condition but shifts the locus of control to the user. Participants explicitly assigned weights to each criterion using interactive sliders (Figure 4). Following any adjustment, the system automatically recalculated the S_j scores, updating the drug rankings and visualizations in real-time to reflect the clinician's specific preferences.

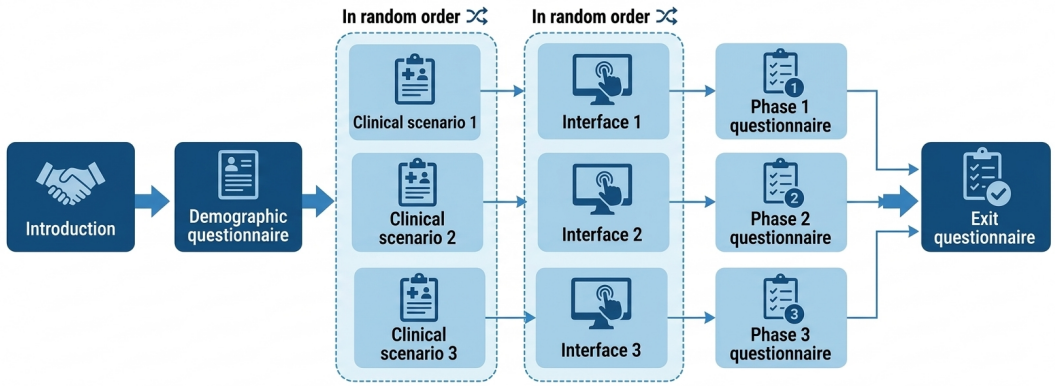

Figure 1: Experimental Design. Participants completed three phases in random order, interacting with the three weighting schemes in a randomized sequence. At each phase, participants selected a treatment before and after interacting with the assigned weighting scheme and completed a utility assessment questionnaire. Finally, an exit questionnaire was administered.

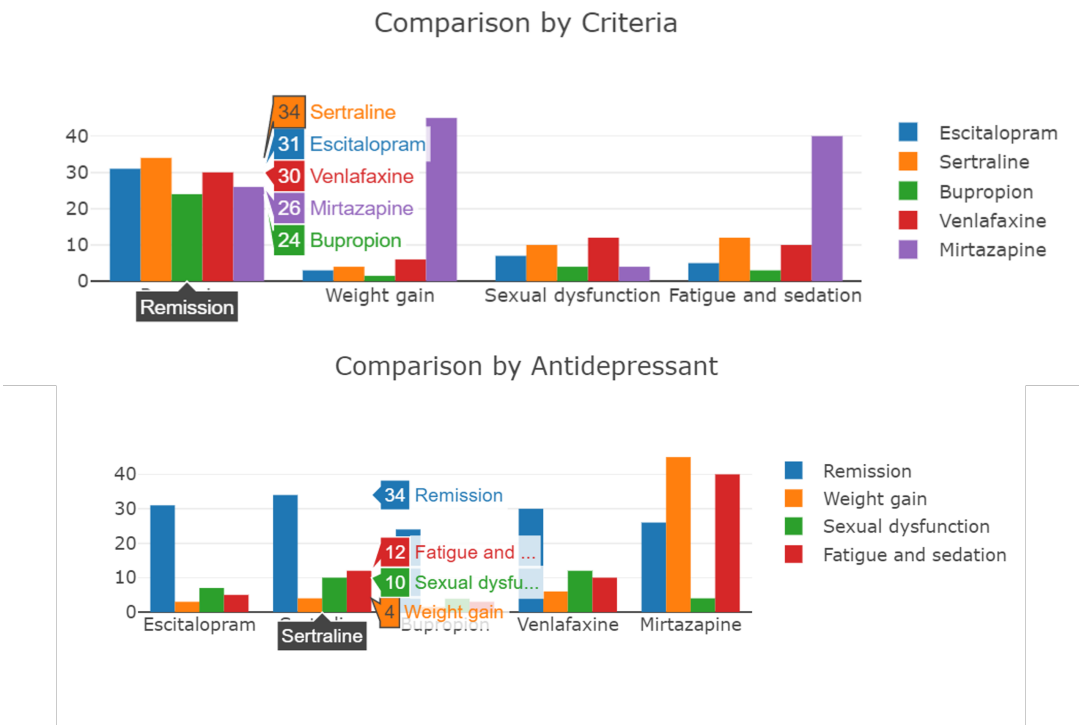

Figure 2: Implicit Weighting. Interactive charts compare drugs by individual outcomes (e.g., remission probability) or by drug profiles. No aggregated score is provided.

2.3 Participants

We recruited 28 clinicians (psychiatrists and primary care physicians / family doctors) from three countries: BLIND FOR REVIEW. Recruitment was conducted via professional networks. Of these, 22 participants (13 psychiatrists, 9 PCPs) completed the full protocol and are included in our analysis. The study was approved by the Institutional Review Board (IRB) of BLIND FOR REVIEW.

2.4 Analysis

We employed a mixed-methods approach combining quantitative outcome measures with qualitative feedback.

Quantitative Analysis: We evaluated four primary metrics:

1. **Perceived Clinical Utility:** Assessed via post-phase questionnaire using a 5-point Likert scale.
2. **Perceived Impact on Decision Making:** Assessed via post-phase questionnaire using a binary item. **Time**

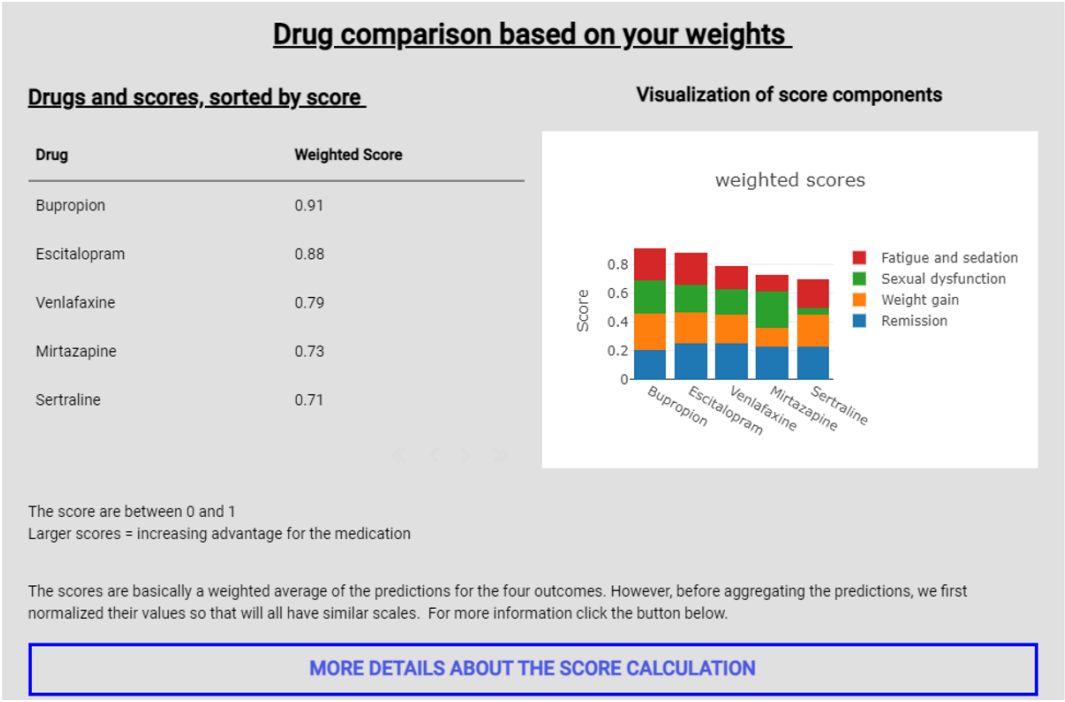

Figure 3: Expert-Derived Weighting. Drugs are ranked by their aggregated S_j scores. Stacked bars provide explainability by showing the contribution of each weighted criterion to the total score.

Spent: We measured the time the participants spent using each of the three schemes.

3. **Observed Impact on Decision Making:** Automatically measured as a binary variable of "treatment change" – this was defined by comparing the participant's initial *unassisted* choice (recorded prior to interacting with the AI-CDSS) with their *final* choice (recorded after interacting with the weighting scheme).

Due to the non-normal distribution of the data and the sample size ($N = 22$), we utilized the non-parametric Kruskal–Wallis test [McKight and Najab, 2010] for main effects across the three conditions, followed by Dunn's test for post-hoc pairwise comparisons.

Qualitative Analysis: To deepen our understanding of user perceptions, participants were offered the opportunity to participate in a semi-structured interview following the experiment, debriefing them on their experience with the three weighting schemes. Three participants (N=3) agreed to take part in this optional phase. Themes were identified using inductive thematic analysis [Khairat *et al.*, 2018] and validated via investigator triangulation [Mehltretter *et al.*, 2019] to ensure reliability in interpreting user attitudes toward clinical utility and agency.

2.5 Clinical Scenarios

The AI-CDSS was applied to three realistic clinical profiles ("Jack", "Emma", and "Sara") representing mild, moderate, and severe depression, respectively (complete details are provided in the Appendix). The probability estimates for each profile (Table 1) were constant across all experimental conditions, maintaining the analytical focus on the weighting scheme and their perceived clinical utility and locus of control.

3 Results

In this section, we present both the quantitative and qualitative results.

3.1 Quantitative Results

Participant Demographics

Our final sample consisted of 22 participants (13 psychiatrists, 9 PCPs) who completed the full protocol. Table 2 details the demographic and professional characteristics of the cohort.

Perceived Clinical Utility

Participants rated the clinical utility of each weighting scheme on a 5-point Likert scale (1=Very Little, 5=Very Much). A significant main effect of the weighting scheme on perceived utility was observed ($H = 10.29, p < 0.01$). Post-hoc analysis indicated that the CDW scheme (Mean: 4.18 ± 0.89) was rated significantly higher than both the Implicit Weighting baseline (Mean: $3.27 \pm 0.86, p < 0.01$) and the EDW scheme (Mean: $3.41 \pm 1.19, p < 0.05$).

Subgroup analysis (Figure 5) showed that psychiatrists significantly preferred the CDW scheme over both alternatives ($H = 8.42, p < 0.05$). PCPs also rated the CDW the highest (Mean: 4.33), though the difference did not reach statistical significance in this subgroup.

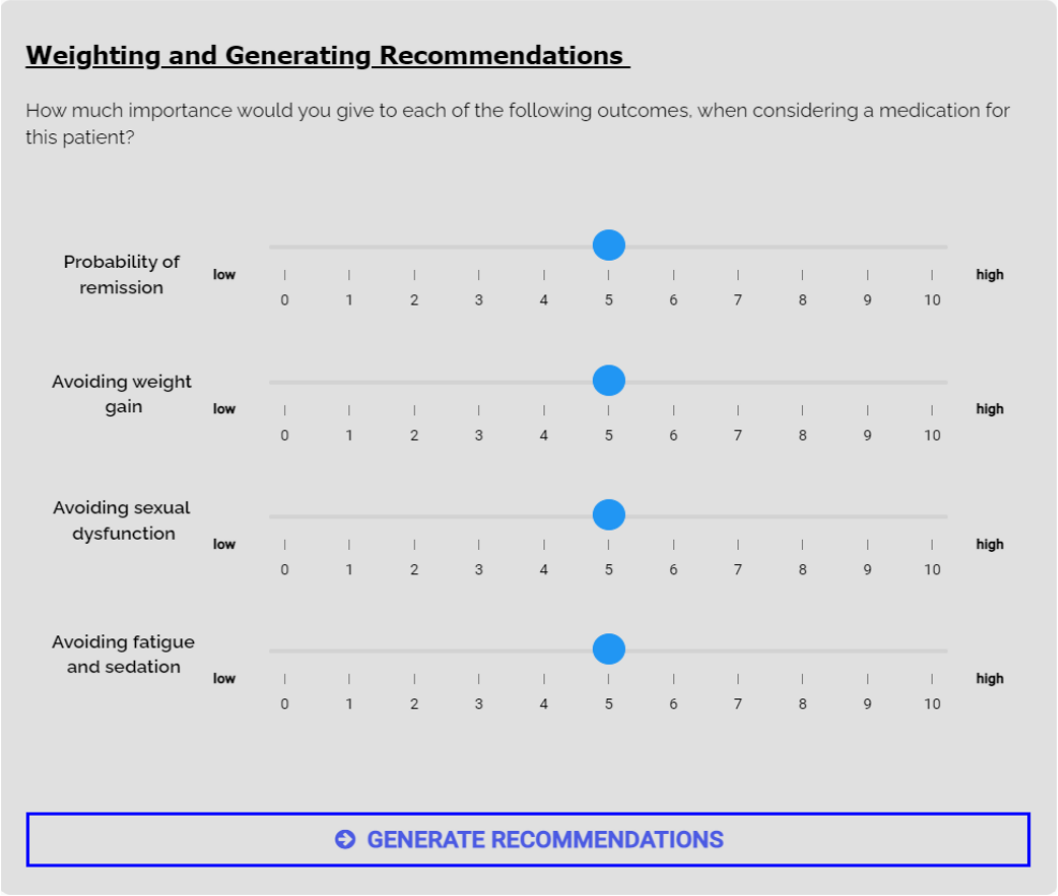

Figure 4: Clinician-Determined Weighting. Drugs are ranked by their S_j scores, which are adjusted dynamically as the user modifies the importance weights via interactive sliders.

Table 1: Predicted probabilities for depression profiles.

Use-case	Medication	Remission	Weight Gain	Sexual Dysf.	Fatigue
Jack (Mild Depression)	Escitalopram	81%	5%	7%	7%
	Sertraline	76%	6%	20%	10%
	Bupropion	69%	1%	3%	8%
	Venlafaxine	79%	8%	8%	16%
	Mirtazapine	75%	20%	1%	22%
Emma (Moderate Depression)	Escitalopram	42%	3%	17%	6%
	Sertraline	46%	5%	25%	11%
	Bupropion	39%	1.2%	5%	3%
	Venlafaxine	44%	6.5%	10%	8%
	Mirtazapine	41%	50%	3%	35%
Sara (Severe Depression)	Escitalopram	31%	3%	7%	5%
	Sertraline	34%	4%	10%	12%
	Bupropion	24%	1.5%	4%	3%
	Venlafaxine	30%	6%	12%	10%
	Mirtazapine	26%	45%	4%	40%

Perceived Impact on Decision Making

Participants self-reported whether the interaction with the AI-CDSS changed their initial decision in each phase (Table 3). While 54.5% reported an affirmative impact in the Implicit Weighting baseline, this was driven largely by PCPs (77.8%) rather than psychiatrists (38.5%). In contrast, the

Table 2: Participant Demographics ($N = 22$).

Attribute	Category	Count (Mean ± SD)
Age	-	47.3 ± 8.53
Gender	Female	11
	Male	11
Experience (Years)	0–4	2
	5–9	6
	10–14	2
	15+	12
Specialty	Psychiatry	13
	Primary Care (PCP)	9
Location	Blind1	10
	Blind2	6
	Blind3	4
	Other	2

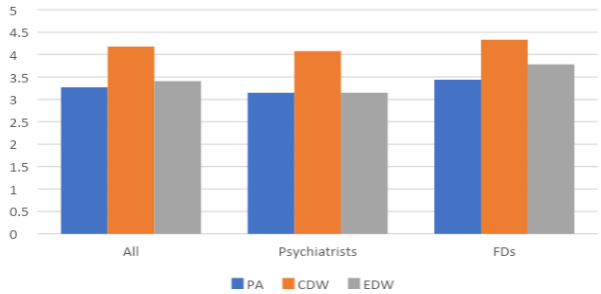

Figure 5: Perceived Clinical Utility. The CDW scheme was rated highest across the full sample (left) and within both specialty subgroups (center, right).

CDW scheme showed a consistent impact across both groups (53.8% of psychiatrists, 66.7% of PCPs). The EDW scheme had the lowest reported impact (36.4% overall), particularly among psychiatrists (23.1%).

Table 3: Perceived Impact on Decision. Percentage of participants indicating that the weighting scheme influenced their treatment choice (affirmative response).

Weighting Scheme	All	Psychiatrists	PCPs
Implicit (Baseline)	54.5%	38.5%	77.8%
EDW	36.4%	23.1%	55.6%
CDW	**59.1%**	**53.8%**	66.7%

Time Spent

While no statistically significant time differences were observed, the highest mean time was spent using the Implicit weighting scheme, with a mean of 404.7 sec (std of 343.8). For comparison, the mean time for using the CDW scheme was 355.8 sec (std of 162.75) and for the EDW scheme 320.6 sec (std of 154.1).

Observed Change in Treatment

Across all 66 trials (22 participants × 3 phases), an initial choice was recorded in 84.8% of cases. Of these, the CDW scheme prompted the highest rate of treatment revision: participants changed their initial choice in **33.3%** of cases. In comparison, the *Implicit* baseline prompted a change in 26.3% of cases, and the EDW scheme in only 15.8% of cases.

3.2 Qualitative Results

Five key themes emerged from manually analyzing the interview transcripts: (1) Clinical Utility & Agency, (2) Integration Within Workflow, and (3) Transparency vs. Automation.

Clinical Utility & Agency: Participants generally favored the CDW for its ability to operationalize their specific clinical intent. As one participant noted: "It made me consider the overall fit of the medication... rather than focusing on one aspect that might not have fit very well". Conversely, the EDW faced resistance, particularly from psychiatrists. One remarked that weighting outcomes is "precisely the clinical work that psychiatrists should be doing," viewing the static weights as an encroachment on their expertise. However, a minority preference for EDW was noted; one PCP preferred the EDW because "it felt like magic" and questioned the point of AI if physicians had to set weights themselves.

Integration Within the Workflow: The CDW scheme was described as a means for *confirmation* and *refinement*. One clinician explained: "It confirmed the decision I had in mind... made me more confident I was not overlooking important information". Another participant stated, "I changed the initial treatment based on the side effects" after seeing the weighted impact.

Transparency vs. Automation: Resistance to the "black box" nature of the EDW scheme was observed. Participants expressed that expert-set weights "are no substitute for engaging patients in their own care," highlighting that the EDW scheme contradicts shared decision-making principles. In contrast, the CDW scheme was seen as a potential bridge for patient communication: "Patients would love to see how their doctor comes to their decision".

4 Discussion

Our results demonstrate a comparatively consistent hierarchy of clinical utility where both specialists and PCPs preferred the CDW scheme over both the Implicit and EDW alternatives.

The favorability of the CDW scheme over the implicit baseline suggests that while clinicians value transparency, they also value algorithmic assistance in aggregating high-dimensional data. The "Dashboard"-like analytics offered by the Implicit weighting scheme, while common, burdens clinicians in our setting with the cognitive load of integrating competing probabilities. The dynamic CDW scheme offloads this computation while retaining the user's strategic intent and agency.

Crucially, the preference for the CDW scheme over the EDW scheme highlights the limitations of "black-box" value alignment, agreeing with a long line of works in the field of human-computer interactions [Loyola-Gonzalez, 2019; Christian, 2020]. Two key mechanisms likely drive this result. First, clinicians, particularly specialists, resisted the static model because it encroached on their core expertise—determining the "value" of a clinical outcome for a

specific patient, aligning with [Khairat *et al.*, 2018; Knop *et al.*, 2022]. By restoring the locus of control to the user, the CDW scheme transformed the AI from a "replacement" to an "assistant", thereby increasing acceptance. Second, as noted in our qualitative feedback, the CDW scheme serves as a boundary object between clinician and patient, allowing for collaborative weighting of side effects - a workflow impossible with rigid expert weights [Elwyn *et al.*, 2012; Kleinerman *et al.*, 2021].

Our findings regarding decision impact, both perceived and observed, further underscore this dynamic. The CDW scheme was not only the most preferred but also the most effective at prompting an AI-informed treatment decision (33.3% change rate). Interestingly, the EDW scheme had the *lowest* impact on decision-making (15.8%), particularly among psychiatrists. This suggests that when AI seemingly "dictates the rules" of the decision (via fixed weights), experts are more likely to dismiss the recommendation entirely and maintain their agency. Conversely, when experts define the rules (via the CDW scheme), they are more willing to accept the AI's predictions (even if these are "black-box"), even when it does not align with their initial intuition.

Finally, the discrepancy between PCPs and psychiatrists reinforces the "Floor-Ceiling" effect in AI-CDSS [Tschandl *et al.*, 2020; Kalyuga *et al.*, 2003]. PCPs, who may have less domain expertise in psychopharmacology, derived high value from both the implicit and CDW schemes. Psychiatrists, conversely, found little value in the implicit weighting scheme but rated the CDW scheme highly. This implies that flexible, controllable AI architectures are essential for engaging domain experts who might otherwise reject decision support [Shulha *et al.*, 2024].

It is important to note that our study has several limitations that offer potentially fruitful avenues for future work. First, the sample size ($N = 22$) limits the statistical power for subgroup analyses, particularly between specialties. While non-parametric tests confirmed main effects, larger cohorts could bring about more nuanced results, such as validating the PCP-Psychiatrist divergence. Second, the study utilized simulated vignettes rather than real patient interactions. While this controlled for case variance, it cannot fully capture the pressures of a live clinical environment or the direct input of a patient in the loop. Third, our implementation of the EDW scheme relied on a single set of three experts. It is thus plausible that a more comprehensive expert evaluation could bring about more acceptable expert-based weights. Finally, both the CDW and EDW schemes assume linear independence between criteria, which is clearly a simplification of reality. While transparent and easily explainable, future iterations could explore non-linear aggregation methods as alternatives.

5 Conclusion

As AI-CDSS move from predictive accuracy to clinical integration, the challenge shifts from *what* the model predicts to "how" those predictions are aggregated. This study demonstrates that for complex multi-criteria CDSSs, like those supporting antidepressant selection, the aggregation scheme could be critical. We found that an algorithmic weighting scheme that prioritizes clinical agency significantly outperforms both implicit weighting (i.e., dashboard-like) and expert-derived weighting in terms of clinical utility and decision impact.

Taken jointly, our findings suggest a design imperative for future "AI for Health" systems: effective AI-powered decision support could greatly benefit from decoupling risk prediction (the AI's job) from value alignment (the clinician's job). By providing transparent, controllable mechanisms for clinicians to explicitly consider trade-offs, we can build systems that not only reduce cognitive load but also respect and enhance clinical agency.

A Appendix: Clinical Scenarios

A.1 Case 1: Jack (Mild-Moderate Depression)

Profile: 82-year-old male, retired, lives alone (widowed/separated). Generally active social life (senior's club) but withdrawing recently. **Medical History:** Hypertension and Type II Diabetes (well-controlled). No prior psychiatric history. **Presenting Symptoms:**

- **Mood:** "Feeling sad", "not myself", loss of enjoyment in hobbies (cards).
- **Anxiety:** Recent panic attacks during decision-making; observed fidgeting/tremors.
- **Vegetative:** Difficulty falling asleep, early morning awakening, increased fatigue.
- **Physical:** Dizziness upon standing (orthostasis), occasional skin itchiness.

Preferences/Concerns:

- Low concern for sexual side effects ("not active anymore").
- High concern for weight gain (diabetes risk).
- Generally adherent to medications (takes BP pills without issue).

Diagnosis: Mild-Moderate Depression. No suicidality.

A.2 Case 2: Emma (Moderate Depression)

Profile: 40-year-old female, lawyer/consultant, married. High-functioning professional facing recent work stress. **Medical History:** None. **Presenting Symptoms:**

- **Mood:** Irritable, sad, feeling guilty ("letting people down"), sobbing during interview.
- **Cognitive:** Difficulty concentrating on reading/work.
- **Suicidality:** Passive ideation ("better if I were dead") but no intent/plan.
- **Vegetative:** Middle insomnia, weight loss (2 lbs), fatigue.
- **Somatic:** Health anxiety (fear of cancer), constipation.

Preferences/Concerns:

- **High concern for sexual dysfunction** (marital strain).
- High concern for weight gain.

- Anxious about side effects in general (medication naïve).

Diagnosis: Moderate-Severe Depression.

A.3 Case 3: Sara (Severe Depression)

Profile: 51-year-old female, unemployed (lost job in food service due to symptoms), lives alone. **Medical History:** None known. **Presenting Symptoms:**

- **Mood:** Depressed mood over 50% of time, significant anhedonia (unable to work).
- **Psychomotor:** Significant slowing (latency in speech).
- **Cognitive:** Severe indecisiveness, inability to manage finances.
- **Suicidality:** Frequent suicidal ideation (several times a week), wishes she were dead, but no active plan.
- **Vegetative:** Severe insomnia (initial, middle, and late), weight loss (5 lbs).

Preferences/Concerns:

- Difficulty articulating specific concerns due to severity of illness.
- General worry about side effects; requires prompting.

Diagnosis: Severe Depression.

References

[Abu-Mahfouz *et al.*, 2026] Mohammad S Abu-Mahfouz, Sarah AlFehaid, Hala M Burqan, and Rabie Adel El Arab. Artificial intelligence in mental health care: a scoping review of reviews. *Frontiers in Psychiatry*, 17:1688043, 2026.

[Aljohani, 2025] Abeer Aljohani. Ai-driven decision-making for personalized elderly care: a fuzzy mcdm-based framework for enhancing treatment recommendations. *BMC Medical Informatics and Decision Making*, 25(1):1–16, 2025.

[Archibald, 2016] Mandy M Archibald. Investigator triangulation: A collaborative strategy with potential for mixed methods research. *Journal of mixed methods research*, 10(3):228–250, 2016.

[Aruldoss *et al.*, 2013] Martin Aruldoss, T Miranda Lakshmi, and V Prasanna Venkatesan. A survey on multi criteria decision making methods and its applications. *American Journal of Information Systems*, 1(1):31–43, 2013.

[Ashton *et al.*, 2005] Adam Keller Ashton, Brenda D Jamerson, Wendy L Weinstein, and Christine Wagoner. Antidepressant-related adverse effects impacting treatment compliance: results of a patient survey. *Current therapeutic research*, 66(2):96–106, 2005.

[Benrimoh *et al.*, 2021] David Benrimoh, Myriam Tanguay-Sela, Kelly Perlman, Sonia Israel, Joseph Mehltretter, Caitrin Armstrong, Robert Fratila, Sagar V Parikh, Jordan F Karp, Katherine Heller, 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(1), 2021.

[Bet *et al.*, 2013] Pierre M Bet, Jacqueline G Hugtenburg, Brenda WJH Penninx, and Witte JG Hoogendijk. Side effects of antidepressants during long-term use in a naturalistic setting. *European Neuropsychopharmacology*, 23(11):1443–1451, 2013.

[Burra *et al.*, 2007] Tara A Burra, Edmund Chen, Roger S McIntyre, Sherry L Grace, Emma Robertson Blackmore, and Donna E Stewart. Predictors of self-reported antidepressant adherence. *Behavioral Medicine*, 32(4):127–134, 2007.

[Charness *et al.*, 2012] Gary Charness, Uri Gneezy, and Michael A Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1):1–8, 2012.

[Christian, 2020] Brian Christian. *The alignment problem: Machine learning and human values*. WW Norton & Company, 2020.

[Dunn, 1964] Olive Jean Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.

[Elhaddad and Hamam, 2024] Malek Elhaddad and Sara Hamam. Ai-driven clinical decision support systems: an ongoing pursuit of potential. *Cureus*, 16(4), 2024.

[Elwyn *et al.*, 2012] Glyn Elwyn, Dominick Frosch, Richard Thomson, Natalie Joseph-Williams, Amy Lloyd, Paul Kinnersley, Emma Cording, Dave Tomson, Carole Dodd, Stephen Rollnick, et al. Shared decision making: a model for clinical practice. *Journal of general internal medicine*, 27(10):1361–1367, 2012.

[Hunot *et al.*, 2007] Vivien M Hunot, Rob Horne, Morven N Leese, and Rachel C Churchill. A cohort study of adherence to antidepressants in primary care: the influence of antidepressant concerns and treatment preferences. *Primary care companion to the Journal of clinical psychiatry*, 9(2):91, 2007.

[Kalyuga *et al.*, 2003] Slava Kalyuga, Paul Ayres, Paul Chandler, and John Sweller. The expertise reversal effect. *Educational psychologist*, 38(1):23–31, 2003.

[Khairat *et al.*, 2018] Saif Khairat, David Marc, William Crosby, Ali Al Sanousi, et al. Reasons for physicians not adopting clinical decision support systems: critical analysis. *JMIR medical informatics*, 6(2):e8912, 2018.

[Kleinerman *et al.*, 2021] Akiva Kleinerman, Ariel Rosenfeld, David Benrimoh, Robert Fratila, Caitrin Armstrong, Joseph Mehltretter, Eliyahu Shneider, Amit Yaniv-Rosenfeld, Jordan Karp, Charles F Reynolds, et al. Treatment selection using prototyping in latent-space with application to depression treatment. *PloS one*, 16(11):e0258400, 2021.

[Knop *et al.*, 2022] Michael Knop, Sebastian Weber, Marius Mueller, and Bjoern Niehaves. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence–enabled clinical decision support systems: literature review. *JMIR Human Factors*, 9(1):e28639, 2022.

[Lam, 2018] Raymond W. Lam. Comparison of 21 antidepressants. *Focus*, 16(4):4s–4s, 2018.

[Loyola-Gonzalez, 2019] Octavio Loyola-Gonzalez. Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE access*, 7:154096–154113, 2019.

[Martin *et al.*, 2022] Vincent P Martin, Jean-Luc Rouas, Pierre Philip, Pierre Fourneret, Jean-Arthur Micoulaud-Franchi, and Christophe Gauld. How does comparison with artificial intelligence shed light on the way clinicians reason? a cross-talk perspective. *Frontiers in Psychiatry*, 13:926286, 2022.

[McKight and Najab, 2010] Patrick E McKight and Julius Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.

[Mehltretter *et al.*, 2019] Joseph Mehltretter, Robert Fratila, David Benrimoh, Adam Kapelner, Kelly Perlman, Emily Snook, Sonia Israel, Marc Miresco, and Gustavo Turecki. Differential treatment benefit prediction for treatment selection in depression: a deep learning analysis of star* d and co-med data. *BioRxiv*, page 679779, 2019.

[Najafi *et al.*, 2023] Amirhossein Najafi, Alireza Nemati, Mahdi Ashrafzadeh, and Sarfaraz Hashemkhani Zolfani. Multiple-criteria decision making, feature selection, and deep learning: a golden triangle for heart disease identification. *Engineering Applications of Artificial Intelligence*, 125:106662, 2023.

[Popescu *et al.*, 2021] Christina Popescu, Grace Golden, David Benrimoh, Myriam Tanguay-Sela, Dominique Slowey, Eryn Lundrigan, Jérôme Williams, Bennet Desormeau, Divyesh Kardani, Tamara Perez, 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(10):e31862, 2021.

[Sarantopoulos *et al.*, 2024] Andreas Sarantopoulos, Christina Mastori Kourmpani, Atshaya Lily Yokarasa, Chiedza Makamanzi, Polyna Antoniou, Nikolaos Spernovasilis, and Constantinos Tsioutis. Artificial intelligence in infectious disease clinical practice: an overview of gaps, opportunities, and limitations. *Tropical Medicine and Infectious Disease*, 9(10):228, 2024.

[Serretti *et al.*, 2010] Alessandro Serretti, Laura Mandelli, M Laura, et al. Antidepressants and body weight: a comprehensive review and meta-analysis. *The Journal of clinical psychiatry*, 71(10):979, 2010.

[Shulha *et al.*, 2024] Michael Shulha, Jordan Hovdebo, Vinita D'Souza, Francis Thibault, and Rola Harmouche. Integrating explainable machine learning in clinical decision support systems: study involving a modified design thinking approach. *JMIR Formative Research*, 8(1):e50475, 2024.

[Thokala *et al.*, 2016] Praveen Thokala, Nancy Devlin, Kevin Marsh, Rob Baltussen, Meindert Boysen, Zoltan Kalo, Thomas Longrenn, Filip Mussen, Stuart Peacock, John Watkins, et al. Multiple criteria decision analysis for health care decision making—an introduction: report 1 of the ispor mcda emerging good practices task force. *Value in health*, 19(1):1–13, 2016.

[Triantaphyllou, 2000] Evangelos Triantaphyllou. Multi-criteria decision making methods. In *Multi-criteria decision making methods: A comparative study*, pages 5–21. Springer, 2000.

[Trivedi *et al.*, 2004] Madhukar H Trivedi, Janet K Kern, Bruce D Grannemann, Kenneth Z Altshuler, and Prabha Sunderajan. A computerized clinical decision support system as a means of implementing depression guidelines. *Psychiatric Services*, 55(8):879–885, 2004.

[Tschandl *et al.*, 2020] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, et al. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, 2020.

[Van Booven *et al.*, 2025] Derek Van Booven, Chen Cheng-Bang, and Manoharan Meenakshy. Limitations of artificial intelligence in healthcare. In *Artificial intelligence in urologic malignancies*, pages 231–246. Elsevier, 2025.

[Zhou *et al.*, 2022] Sijia Zhou, Jingping Zhao, and Lulu Zhang. Application of artificial intelligence on psychological interventions and diagnosis: An overview. *Frontiers in Psychiatry*, 13:811665, 2022.

[Zionts, 1979] Stanley Zionts. Mcdm—if not a roman numeral, then what? *Interfaces*, 9(4):94–101, 1979.