Research Paper

A mixed-methods feasibility study of a novel AI-enabled, web-based, clinical decision support system for the treatment of major depression in adults

Sabrina Qassim [a,c], Grace Golden [a,c], Dominique Slowey [b,c], Mary Sarfas [b,c], Kate Whitmore [b,c], Tamara Perez [b,c], Elizabeth Strong [b,c], Eryn Lundrigan [b,c], Marie-Jeanne Fradette [b,c], Jacob Baxter [b,c], Bennet Desormeau [b,c], Myriam Tanguay-Sela [c], Christina Popescu [c], Sonia Israel [c], Kelly Perlman [b,c], Caitrin Armstrong [c], Robert Fratila [c], Joseph Mehltretter [c], Karl Looper [i], Warren Steiner [i], Soham Rej [i], Jordan F. Karp [d], Katherine Heller [e], Sagar V. Parikh [f], Rebecca McGuire-Snieckus [g], Manuela Ferrari [h], Howard Margolese [i], David Benrimoh [c,i,*]

[a] University of Waterloo, Canada, ON, Waterloo, N2L 3G1
[b] McGill University, Canada, Quebec, Montreal, H3A 0G4
[c] Aifred Health Inc., Canada, Quebec, Montreal, H3J 1M1
[d] University of Arizona, United States, Arizona, Tucson, 85721
[e] Duke University, United States, North Carolina, Durham, 27708
[f] University of Michigan, United States, Ann Arbor, 48109
[g] Barts and the London School of Medicine, United Kingdom, England, London, E1 2AD
[h] Douglas Mental Health University Institute, McGill University, Canada, Quebec, Verdun, H4H 1R3
[i] Department of Psychiatry, McGill University, Canada, Quebec, Montreal, H3A 0G4

ARTICLE INFO

ABSTRACT

Background: The objective of this paper is to discuss perceived clinical utility and impact on physician-patient relationship of a novel, artificial-intelligence (AI) enabled clinical decision support system (CDSS) for use in treating adults with major depression.

Methods: A single arm, naturalistic follow-up study aimed at assessing the acceptability and useability of the software. Patients had a baseline appointment, followed by a minimum of two appointments with the CDSS. Study exit questionnaires and interviews were conducted to assess perceived clinical utility, impact on patient-physician relationship, and understanding and trust. 7 physicians and 17 patients, of which 14 completed, consented to participate.

Results: 86 % of physicians (6/7) felt the information provided by the CDSS provided more comprehensive understanding of patient situations. 71 % (5/7) felt the information was helpful. 86 % of physicians (6/7) reported the AI/predictive model was useful when deciding treatment. 62 % of patients (8/13) reported improved care due to the tool, and 46 %(6/13) reported a significantly or somewhat improved physician-patient relationship 54 % reported no change. 71 % of physicians (5/7) and 62 % of patients (8/13) rated trusting the tool.

Limitations: Small sample size and treatment changes prior to CDSS introduction limits ability to verify impact on outcomes.

Conclusions: Qualitative results from 12 patient exit interviews are analyzed and presented. Findings suggest physicians perceived the tool as useful in conducting appointments and used it while deciding treatment. Physicians and patients generally found the tool trustworthy, and it may have positive effects on physician-patient relationships. (Study identifier: NCT04061642).

* Corresponding author at: 304-1600 Rue Notre-Dame O, Montreal, QC, Canada, H3J 1M1.
  E-mail address: david.benrimoh@aifredhealth.com (D. Benrimoh).

1. Introduction

Clinical Decision Support Systems (CDSS; Sim et al., 2001) in mental health treatment, including systems that incorporate artificial intelligence (AI) to assist with treatment management, are being developed and researched at an increasing rate (Tran et al., 2019). CDSS are tools that serve to assist clinicians in making accurate and informed decisions regarding treatment and are meant to supplement the clinician-patient relationship (Patel et al., 2008). In this paper we discuss Aifred, a novel CDSS that aims to support clinicians and patients as they make choices about the treatment of major depression.

The Aifred CDSS is described in detail in previous work (Benrimoh et al., 2020; Popescu et al., 2021). The Aifred platform is an AI-enabled clinical decision support system intended to supplement physician decision making and support shared decision making in the selecting and management of treatments for Major Depressive Disorder (MDD) in adults. The prediction provided by the AI system are the results of deep learning machine learning model trained on clinical and demographic data which is then used to provide personalized probability of remission for specific treatment options (see Mehltretter et al., 2020 for technical details). Clinicians have access to answers and trends from patient self-report questionnaires which they can track assisting them in the practive of measurement-based care (MBC; Goldberg et al., 2018; Scott and Lewis, 2015). On the platform they can access a clinical algorithm, based on the CANMANT 2016 guidelines for depression treatment (Kennedy et al., 2016), which provides step-by-step information about treatment management. When it is time to select a treatment, the AI provides remission probabilities for first line treatments. The rule-based clinical algorithm then tracks depression scores from baseline to subsequent visits to determine whether a response to medication has been observed and provides relevant information from the guidelines about next appropriate steps. In this way, the AI helps with initial treatment selection, and the clinical algorithm assists with treatment management. Patients have access to their questionnaire scores over time, allowing them to have information about their symptoms over time, as well as their active and past treatments. As this platform is intended as decision support, the physician is always in control of clinical decisions and may choose not act in accordance with either the guidelines or the AI-based predictions. More details can be found in (Benrimoh et al., 2020; Popescu et al., 2021).

Finally, the Aifred CDSS incorporates a novel element: an artificial intelligence (AI)-based predictive model. This model inputs patient clinical and demographic characteristics to generate treatment-specific remission probabilities for a number of commonly prescribed antidepressants. Pharmaceutical treatments are presented to the physician alongside the remission probability for each medication for which the model has data and other guideline-derived clinical information, such as dosing or specific symptoms for which the treatment may have evidence of effectiveness (e.g., fatigue, sleep or pain) (Kennedy et al., 2016). Clinicians also had access to both baseline population remission rates based on data that was used to train the CDSS as well as an "interpretability report" which consisted of the up to five patient variables which had the greatest impact in the prediction for each medication (see Mehltretter et al., 2020, for an example). The CDSS is designed to be used during sessions with patients and to assist in shared decision-making (Benrimoh et al., 2020). The rationale for including the AI component is that personalization of treatment in depression remains a challenge, often relying on trial and error. As such, a point-of-care predictive model may enable identifying optimally effective treatments for patients earlier in their care (Benrimoh et al., 2018).

The Aifred CDSS, or simply "tool", has previously been shown to be easy to use by clinicians and perceived as clinically useful and trustworthy in a simulation center study (Benrimoh et al., 2020; Tanguay-Sela et al., 2022). The current paper reports on the results of an in-clinic Feasibility study of the Aifred CDSS. The primary outcome of this study was clinical feasibility, as demonstrated by no increase in appointment length (Popescu et al., 2021). This feasibility study had as its primary objective to establish the feasibility of this tool in clinical practice and to assess impact on the clinician-patient relationship. In addition, we sought to explore perceptions of its clinical utility by clinicians and patients. Clinicians and patients are not likely to use tools which pose barriers in terms of feasibility, and which are not perceived to be clinically useful. As such, given the expense of clinical trials and the requirement to expose large numbers of patients to new tools during a trial, we argue that it is important to establish the feasibility and perceived utility of tools such as the one described here prior to their incorporation into clinical trials aimed at establishing tool effectiveness in order to ensure that the tool is ready for large-scale testing.

In previous work (Popescu et al., 2021) we have focused on the quantitative aspects of this study, in particular the demonstration that appointment lengths did not increase with tool introduction which was a key element of establishing tool feasibility based on comments from clinicians during the tool's construction. In the current article, we focus on the qualitative results of this study, detailing the perceptions of clinicians and patients who used the tool during the study. This paper considers both quantitative and previously unreported qualitative results from this study, with a focus on discussing perceived clinical utility, trust, and impacts on the clinician-patient relationship of the Aifred CDSS. We note that this tool is an investigational device not yet labeled for use and that this study was carried out under investigational testing authorization by Health Canada.

2. Methods

Ethics approval for the study was obtained by the Research Ethics Boards of the Douglas Mental Health University Institute (clinicaltrials.gov identifier: NCT04061642). The study was carried out according to the Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans (Canadian Institutes of Health Research, 2005), with participants providing written informed consent.

The study was a single arm, naturalistic follow-up study aimed at assessing the acceptability and useability of the software, conducted from January to November 2020. The study sample consisted of two populations: 1) physicians, including psychiatrists and family physicians and 2) patients of these physicians. Patients were eligible if they were over the age of 18, capable of providing informed consent, diagnosed by their physician with an active major depression using DSM-5 criteria (APA, 2013) which the clinician felt was the primary mental disorder being treated, and did not have bipolar disorder. 17 patients consented to participate in the study, of which 14 completed, and qualitative data was obtained for 12 participants. Detailed methods, including discussion of sample size, can be found in Popescu et al., 2021.

The study was conducted at university hospitals, primary care, and psychiatric clinics in Québec, Canada. Participating physicians were invited to an introductory session to learn how the AI model was created and trained on how to use the CDSS. Patients who met the eligibility criteria and provided informed consent had accounts created that were linked to their respective physicians. Due to COVID-19 and public health recommendations, the study was adapted to be completed entirely via telemedicine early in the course of the study.

After patients provided informed consent, their next appointment with their physician served as the baseline appointment. It is important to note that the tool was not used by the physician during this appointment. This was because the primary outcome for the study was change in appointment length, and this design allowed the measurement of appointment length pre- and post-CDSS introduction. Patients were then seen at least two subsequent appointments, termed visit 1 and visit 2, during which the tool was used by the physician. At baseline, patients completed a number of questionnaires including a demographics questionnaire, a Patient Health Questionnaire-9 (PHQ-9) to track depressive symptoms and severity over time, a General Anxiety Disorder-7 (GAD-7) to track anxiety symptoms and severity over time, an Alcohol Use

Disorders Identification Test (AUDIT) to screen for potential harmful use of alcohol, a Drug Abuse Screen Test-10 (DAST-10) to screen for potential problematic drug use, a Standardized Assessment of Personality-Abbreviated Scale Self-Assessment (SAPAS-SA) to screen for potential personality disorders, a World Health Organization Disability Assessment Schedule (WHODAS) to assess current disability, a Quick Inventory of Depressive Symptomatology (QIDS) to report symptoms of depression, an Adverse Childhood Experiences (ACE) questionnaire to assess adverse events that occurred in childhood, and an Inventory of Depressive Symptomatology, Clinician Rating (IDS-C), a researcher-assessed questionnaire about depressive symptoms (Kroenke et al., 2001; Spitzer et al., 2006; Bohn et al., 1995; Skinner, 1982; Merlhiot, 2014; Üstün et al., 2010; Yeung et al., 2012; Felitti et al., 1998; Rush et al., 2000). Subsequently, on a weekly basis patients were asked to complete the PHQ-9 and GAD-7 to assess depressive and anxious symptoms, as well as the Patient-Rated Inventory of Side Effects-20 (PRISE-20) and Frequency, Intensity, and Burden of Side Effects Rating (FIBSER), which, together, allow for the assessment of the presence, frequency, intensity, and burden of medication side effects (Katz et al., 2012; Wisniewski et al., 2006). Patients were reminded to complete questionnaires by automated email notifications.

Following visit 2, patients were administered the Scale to Assess Therapeutic Relationships in Community Mental Health Care - Patient Version (STAR-P) to study the quality of the patient-physician relationship as perceived by each patient (McGuire-Snieckus et al., 2007), as well as a customized exit questionnaire and custom semi-structured interview to explore the patient experience with the tool using a quantitative and qualitative approach, respectively. After all of a physician's patients had completed visit 2, they were similarly administered the Scale to Assess Therapeutic Relationships in Community Mental Health Care - Clinician Version (STAR-C) (McGuire-Snieckus et al., 2007), as well as a customized exit questionnaire and custom semi-structured interview, all of which explored concepts akin to the patient versions but from a clinician's perspective. During all interviews, research assistants (RAs) took extensive written notes and these formed the dataset for the qualitative analysis; RAs were instructed to take verbatim notes as recording and transcription was not possible because of resource limitations. Clinicians and patients were compensated for their time. The study flow was documented in Popescu et al., 2021.

2.1. Analysis

2.1.1. Quantitative

Descriptive statistics are presented for the demographics questionnaire, baseline clinical questionnaires, STAR and customized exit questionnaire. Quantitative results regarding main study outcome (appointment length) can be found in Popescu et al., 2021.

2.1.2. Qualitative analysis

The purpose of the qualitative analysis was to explore patient and clinician perspectives regarding the regarding their experience using the tool in the study. In particular, we were interested in understanding perceptions of usability, clinical utility, and effects on the clinician-patient relationship. Written notes obtained during the 20–30 min patient and physician semi-structured interviews were reviewed to extract initial themes prior to data analysis, as a means of commencing the codification of data while minimizing bias. RAs met patients and clinicians one-on-one and asked a set of questions, with instructions to record participant responses in writing; follow-up questions were permitted for clarification or to more fully explore a participant's comment. 13 RAs were familiarized with the data and extracted an initial set of categories and themes based on the experiences of patients and physicians in the study. RAs then reviewed the data once more to refine the initial themes extracted, which consisted of collapsing, splitting, eliminating and creating new themes. Five categories were identified: 1) physician; 2) patient; 3) physician-patient relationship; 4) COVID-19 and 5)

miscellaneous. Themes were identified within the categories. Overarching themes encompassed similar concepts, spanning multiple participants, as found within the data. Themes could be further separated into sub-themes. The physician category consisted of 16 themes, the patient category had 15 themes, both the physician-patient relationship and COVID-19 categories had 5 themes and the miscellaneous category had 2 sub-themes (see Supplementary Material for sub-themes). Using an inductive thematic analysis approach and following the steps laid out in Braun and Clark, 2006, data was coded without trying to fit it into a pre-existing coding frame or the researcher's analytic preconceptions, and allowed the themes identified to be strongly linked to the data themselves.

RAs independently read and coded excerpts of the data into subheadings of the thematic table. All RAs were assigned the data corresponding to all participants, so that each participant's data was independently coded by all RAs. This data was then condensed by collapsing some of the themes and rearranging the subheadings to reduce the overlap of excerpts between subheadings. The RAs then independently reread and coded all of the data into a final summary table. This stage ensured that any data that had been missed in earlier coding stages could be added, and also validated the new themes in relation to the full data set. Any discrepancies or disagreements between RA's were resolved through discussion until a consensus was reached. When there was a discrepancy, each RA was allotted time to describe or discuss why they thought a given statement belonged in a given sub-theme; by consensus, we decided what sub-theme was more appropriate for each except of data. Finally, data were arranged into four final categories of interest to the central themes of this paper (clinical utility and clinician-patient relationship). These were Clinical Utility and Useability of the App, Impact on Physician-Patient Interaction and Relationship, Experience of Measurement-Based Care, and Understanding and Trust.

The main qualitative research question was: how do patients and clinicians perceive the clinical utility and trust in the tool, and its impact on the clinician-patient relationship?

3. Results

3.1. Quantitative results

3.1.1. Sample description

Ten physicians consented to participate in the study; however 3 were unable to recruit patients due to interruptions related to COVID-19 in their clinical practice. As such, the final physician sample consisted of 7 physicians (4 general practitioners, 3 psychiatrists). Twenty patients were approached to join the study, with 17 consenting to participate. One patient dropped out before the baseline appointment and 2 dropped out after the baseline appointment but before visit 1 and therefore did not interact with the CDSS and their physician together in a session. This resulted in 14 patients (9 women and 5 men) completing the study (defined as completing visit 2). Mean participant age was 36.43 years (SD 14.84). See table 1 for patient demographics and table 2 for baseline clinical questionnaire scores.

3.2. Descriptive results

3.2.1. Physician-Patient relationship

In order to determine the quality of the patient-physician relationship, physicians ($n = 7$) completed the STAR-C and patients ($n = 13$) completed the STAR-P at the end of the study. The questionnaire was not administered at the beginning of the study given the range of patient-physician relationship qualities and duration that could not be controlled for due to the small sample size. At the end of the study, patients had a mean total score of 42.69 (SD 5.57) and physicians had a mean total score of 40.29 (SD 5.65). In the original STAR study, patients had a mean total score 38.4 (SD 12.0) and physicians had a mean total

Table 1
Patient demographic characteristics.

Patient characteristics	n (%)	Mean (SD)
Age (years; $n = 14$)	14 (N/A)	36.43 (14.84)
Sex ($n = 14$)		
Male	5(35.71)	
Female	9(64.29)	
Gender ($n = 14$)		
Men	5(35.71)	
Women	9(64.29)	
Ethnicity ($n = 13$)		
Caucasian	10(76.92)	
Caribbean	1(7.69)	
African or African American	1(7.69)	
Unanswered	1(7.69)	
Residency Status ($n = 13$)		
Canadian citizen	11(84.62)	
Immigrant status (>5 years ago)	1(7.69)	
Immigrant status (<5 years ago)	1(7.69)	
Adoption Status ($n = 13$)		
Not adopted	12(92.31)	
Adopted	1(7.69)	
Employment Status ($n = 13$)		
Full-Time	7(53.85)	
Part-Time	1(7.69)	
Disability (not working)	2(15.38)	
Unemployed + volunteer work	2(15.38)	
Unemployed	1(7.69)	
Relationship Status ($n = 13$)		
Married	4(30.77)	
Divorced	1(7.69)	
Dating a single partner	2(15.38)	
Not in a relationship	6(46.15)	
Years of Education ($n = 12$)	12(N/A)	15 (6.42)
Income ($ CA) ($n = 10$)	10(N/A)	82,333.33 (70,099.93)

score of 31.5 (SD 6.9). The maximum score for both questionnaires is 48, with higher scores indicating a stronger relationship. The mean STAR-P subscores for positive clinical input, positive collaboration and non-supportive clinical input were 10.23 ± 1.88, 21.92 ± 2.24 and 10.54 ± 2.58, respectively, compared to 9.3 ± 3.0, 19.9 ± 6.7 and 9.3 ± 3.3, the subscores in the original STAR-P. The mean STAR-C subscores for positive clinician input, positive collaboration and emotional difficulties were 9.93 ± 1.38, 17 ± 2.60 and 10.57 ± 1.79, respectively, compared to 8.9 ± 1.6, 7.4 ± 2.7 and 15.3 ± 4.0, the subscores in the original STAR-C. As such, after using the tool there were no marked differences in our sample with respect to clinician-patient relationship compared to a reference population. In addition, the customized exit questionnaire was completed by 13 patients. 46 % of the patients (6/13) felt that the app significantly or somewhat improved the patient-physician relationship and the other 7 (54 %) felt that it did not change.

3.2.2. Perceived clinical utility

86 % of the physicians (6/7) reported that the information provided by the CDSS was useful for getting a better sense of their patients' situations and 71 % (5/7) felt that the information provided by the tool was helpful, indicated by a 4 or 5 on the 5-Point Likert scale. 71 % (5/7) reported that they believed they would continue to use this application after the study and 29 % (2/7) reported that they would not. 57 % of physicians (4/7) reported that they would use the AI/predictive model in future clinical practice, and the other 43 % (3/7) was unsure.

Importantly, 86 % of the physicians (6/7) thought the AI/predictive model was somewhat or very useful in helping them make their treatment decisions, indicated as a 4 or 5 on the 5-Point Likert scale, and one physician was unsure about its usefulness. 71 % (5/7) of the physicians rated that they trusted the tool, indicated as a 4 or 5 on the 5-Point Likert scale.

As indicated by scores of 4 and 5 on the 5 point Likert scale, 62 % of the patients (8/13) rated that they trusted the tool. In addition, 62 % of patients (8/13) felt that the tool allowed them to fully report how they

Table 2
Baseline questionnaire scores. Baseline questionnaires included a GAD-7 (Generalized Anxiety Disorder-7), PHQ-9 (Patient Health Questionnaire-9), SAPAS-SA (Standardized Assessment of Personality-Abbreviated Scale Self-Assessment), DAST-10 (Drug Abuse Screen Test-10), AUDIT (Alcohol Use Disorders Identification Test), QIDS (Quick Inventory of Depressive Symptomatology), WHODAS (World Health Organization Disability Assessment Schedule), ACE (Adverse Childhood Experiences), and an IDS-C (Inventory of Depressive Symptomatology, Clinician Rating).

Questionnaire Baseline Score[1]	Frequency, n (%)	Mean (SD)
GAD-7 ($n = 14$)		12.21 (5.81)
0–4: no or minimal anxiety	1(7.14)	
5–9: mild anxiety	4(28.57)	
10–14: moderate anxiety	3(21.43)	
15–21: severe anxiety	6(42.86)	
PHQ-9 ($n = 15$)		14.80 (5.61)
0–4: minimal or no depression	1(6.67)	
5–9: mild depression	1(6.67)	
10–14: moderate depression	4(26.67)	
15–19: moderately severe depression	7(46.67)	
20–27: severe depression	2(13.33)	
SAPAS-SA ($n = 15$)		3.53 (2.23)
< 3 points: negative screen for personality disorder	5(33.33)	
≥ 3 points: positive screen for personality disorder. Further clinical evaluation is warranted.	10(66.67)	
DAST-10 ($n = 15$)		2.40 (3.60)
0: no problems reported	8(53.33 %)	
1–2: low level	3(20.00)	
3–5: moderate level	0(0.00)	
6–8: substantial level	3(20.00)	
9–10: severe level	1(6.67)	
AUDIT ($n = 15$)		4.40 (3.58)
< 8: negative screen for harmful alcohol use	12(80.00)	
≥ 8: positive screen for harmful alcohol use	3(20.00)	
QIDS ($n = 12$)		13.12 (6.04)
1–5: no depression	2(16.67)	
6–10: mild depression	2(16.67)	
11–15: moderate depression	3(25.00)	
16–20: severe depression	4(33.33)	
21–27: very severe depression	1(8.33)	
WHODAS ($n = 13$)	N/A	54.62 (27.08)
ACE ($n = 13$)		1.62 (1.45)
0 adverse childhood events	4(30.77)	
1 adverse childhood event	3(23.08)	
2 adverse childhood events	1(7.69)	
3 adverse childhood events	4(30.77)	
4+ adverse childhood events	1(7.69)	
IDS-C ($n = 13$)		30.00 (12.88)
0–11: no depression	0(0.00)	
12–23: mild depression	4(30.77)	
24–36: moderate depression	5(38.46)	
37–46: severe depression	3(23.08)	
47–84: very severe depression	1(7.69)	

were feeling and 62 % of patients (8/13) reported that the questions covered the things most important to them. Furthermore, 77 % (10/13) of patients gave scores of 4 and 5 in response to whether keeping track of their symptoms and treatments in the tool helped their physician better manage their care. Overall, 62 % of patients (8/13) felt the app improved their care and the other 38 % (5/13) thought the tool had no effect. No patients noted that their care worsened as a result of the tool.

3.3. Qualitative results

All physicians (7/7) who recruited patients in the study completed a

final interview. In addition, 12 of the 14 patients who completed the study completed an end interview; of these, 7 patients were patients of GPs and 5 were patients of psychiatrists. 2/14 patients could not be contacted to complete a final interview. Below are the data from these interviews, organized into four categories: 1) Clinical utility and use-ability of the application; 2) Physician-patient interaction and relationship; 3) Experience of measurement-based care; and 4) Understanding and trust. Sub-categories are presented in each category. Note that when a certain number of participants are noted as making a statement, this does not imply that the remainder of participants made the opposite statement.

1 Perceived Clinical Utility and Useability of the App

3.3.1. Ease of use

Overall, 10 of 12 patients and 6 of 7 physicians reported that the app was easy to use. Despite noting the tool was easy to use, 3 physicians reported they would have felt more comfortable with the tool with more time and practice. Further comments regarding ease of use can be found in the supplementary material and Box 1.

Box 1

▓.

Physician and patient quotes:

Clinical Utility and Useability of the App
Physicians and patients shared comments regarding ease of using the app, as well as its clinical utility.
Physician 1 reported the tool was "easy and user-friendly, steps and suggestions are well explained."
Physician 2 shared that "reading the change in scores [prior to appointments] definitely gave some orientation as to where the patient is going and if things are improving or not"
Patient 1 felt the tool "allowed [the] doctor to understand symptoms more when she couldn't verbalize how she was feeling."
Patient 2 also noted that they would continue to use the tool in a maintenance capacity, sharing they "would use it but a couple of times per year… [to track] mood changes, [but] not week to week"

Impact on Physician-Patient Interaction and Relationship
Physicians and patients agree that the tool improved patient-clinician interactions.
Patient 3 reported feeling the tool "would be very useful for a new patient-physician relationship or a doctor without much time to give them info on when they're not doing well and to take them seriously."
Physician 3 stated they "felt like there was another outside kind of help and of objectification and that helpful the rapport with the patient"
The tool also improved patient engagement in their own care.
Multiple patients reported they were "better able to self-reflect and articulate how [they were] feeling at visits"

Measurement-Based Care: Patient Perspective
Patients reported the tool brought more awareness to their changing thoughts and feelings.
Patient 2 shared they were able to "reflect on how [their] past week has been [and] helped her remember her moods."
Patient 3 stated "mood really fluctuates during depression so I think it really helped to point out that there were days where I felt ok,"
Patient 4 shared "there's actually something nice about filling in these questionnaires more often. The questions always feel a little off the first time so doing it multiple times takes away the stress of having something riding on me to do it right. […] [I] like that this is a neutral way to be asked about your feelings without having a person or putting on a front for others (such as friends). I enjoyed it, filling in the questionnaires was a useful therapy."
Patient 5 reported: "I know how I feel. I know. And I'm fed up with it; It makes me realize certain things that I was ignoring. When you're like this, you don't think about what you feel. You know you're different, but this helps you realize how you are; It helped me to realize what it is from the questions, and make me realize what I have. I don't know about depression. You hear about people with depression, but you don't know anything about it until you have it. Even the people at work knew because they saw I was not the same, but I did not realize it."

Understanding and Trust
Physicians generally reported trusting the tool enough to integrate it into their care.
Physician 4 shared the tool suggestions "were reasonable and mirrored guidelines"
Patients shared little concern regarding their understanding of the tool.
Patient 4 reported "As long as it works, it works – whether it's magic or AI"

3.3.2. Perceived clinical utility

With respect to clinical utility as perceived by patients, some patients were able to identify how their questionnaire responses were being used by their physician, with one patient expressing that if they "reported higher anxiety levels (for example) that week, [her] doctor saw that in [the] appointment when [they] looked at tool so [they] increase[d] medication." 3 of 12 patients (2 patients of GPs; 1 patient of a psychiatrist) commented that the tool did not necessarily have an effect on their care as they had already been on medication. This was expected as part of the study design, where the tool was introduced after the baseline visit (and therefore after pharmacotherapy was likely to have been initiated).

Patients reported how they used questionnaire responses and seeing their trends graphed to facilitate better communication between themselves and their care provider. One patient shared that it helped in "highlighting my priorities". Another patient (of a GP) stated that "I think what it accomplished was to give me the info I needed to transmit to my GP when I spoke to him. I could look back and reflect on how I felt. That info being there in black and white for me made it easier to be prepared to speak to him." Another patient (of a GP) discussed how the tool "allowed [the] doctor to understand symptoms more when she couldn't verbalize how she was feeling."

Physicians commented on the clinical utility of the CDSS, with particular emphasis on how they interpreted and used the tool's clinical algorithm. One doctor described that it was "nice to objectify scores, as [they] were able to explain what [they] were doing, [and] told the patient what the algorithm suggested". Another physician described their treatment decisions using the tool in the context of the CANMAT guidelines, saying that it is "helpful to know that clinical decisions were supported by CANMAT guidelines, not just physician recommendations and/or opinions". Interestingly, 4/7 (2 Psychiatrists, 2 GPs) doctors shared comments regarding their openness to prescribe medications they may not otherwise have considered based on the information presented within the algorithm module.

A trend that emerged from physicians is that the clinical algorithm and AI allowed doctors to reflect on or confirm their decision-making. One doctor gave an example of this, wherein they reviewed their decision based on the AI, "[I] looked at numbers and trends, [and the AI] gave me examples of why patients were better, worse, or the same". Similarly, 2 physicians, 1 psychiatrist and 1 GP, reported that they would consider the AI recommendations if patients weren't responding to current treatment; one psychiatrist stated "If the patient didn't get better on current meds, then I would change the medication and look at what the algo said. Results were interesting, I haven't used it a lot in practice but if my patient didn't get well, then I would have use[d] the tool more", though they also suggested the AI follow patients for a longer time (6 months) prior to making recommendations. Despite positive comments that the AI and CDSS provided "useful info, [and] remind[s] you treatment only has a chance of working. Maybe how long to try medication for before reassessing," one psychiatrist also noted that when the difference in predicted remission between treatments was low that they were less likely to take this into account during decision making (as they put it, "on a population level, 5 % [improvement in predicted remission] can make a big difference, but not for the person you're treating that specific day"). Using the CDSS as a confirmation of clinical decisions was noted by 5/7 physicians (1 psychiatrist, 4 GPs). No physicians reported negative effects on patient outcomes as a result of using the tool.

One key aspect of assessing perceived clinical utility as well as barriers to use is considering whether or not users are willing to continue using the application in the future. 9 of 12 patients (6 patients of GPs; 3 patients of psychiatrists) reported that they would, as one patient put it, "like to continue using [the app] to monitor" how they are feeling. 2 patients reported they would use the app if there were changes to their treatment plan. 5 of 7 physicians said they were willing to continue using the tool after the study (2 GPs; 3 psychiatrists).

1 Impact on Physician-Patient Interaction and Relationship

All patients and clinicians noted neutral or positive impacts of the application on the physician-patient relationship. One psychiatrist shared that the app may have been beneficial in their doctor-patient relationship stating, "I felt like there was another outside kind of help and of objectification [which facilitated] the rapport with the patient".

4 patients shared that the app in some way, as one patient put it, "enhanced" their relationship with their physician. One patient noted that the tool helped to facilitate conversations: "[the app] made it easier, as sometimes I have trouble verbalizing how I am feeling, so it helped the relationship." The 3 other patients cited their increased involvement and understanding as the main factor for this change. One patient shared, "[using the app] made things a lot easier, […] not many of my other doctors have been able to make me feel good about treatment [and the app] definitely [had] a positive impact". Similarly, a patient shared that "tracking symptoms/treatments helped [the] physician better manage patient care," while another commented, "[the app] improved [the relationship by introducing] more knowledge and understanding in the relationship." Other physicians and patients reported neutral impacts on the relationship.

There were comments supporting the use of the tool in facilitating shared decision making. For example, a clinician asserted that "it's useful to have the feedback that the patient is improving or not. When I looked at the scales and saw that there was no improvement, I could speak about it with the patient and use that in my discussion with them". 2 patients commented that they enjoyed the shared approach saying that they "appreciated that the doctor included [them] in app-use appointments" and that the app "improve[d] care on how they dealt with [their] depression together".

3.3.3. Patient agency and self-management

Two physicians reported feeling that the app improves patient agency due to increased involvement required by questionnaires and the symptom tracking features of the app, with one physician stating, "[the app] improved things [by making] patient[s] more active participants [in their care and allowing them to] see objective changes on questionnaires". Another physician expressed that the tool "made [the patient] more accountable" while another found the app to be "[of] benefit [as it] allowed patients to reflect on symptoms and see what type of questions others in their situation were being asked." Mirroring this sentiment, the 9 patients who commented on patient agency all reported feeling that the app improved their agency and communication. For example, 1 patient credited the app with enabling him to be, "better able to self-reflect and articulate how he was feeling at visits," a sentiment he shared with 3 additional patients. These patients communicated that the app gave them the opportunity to divulge certain information they might not have otherwise shared because they didn't know it was relevant, because they found the experience difficult to express in words, or because they "don't feel comfortable expressing [these experiences] out loud".

1 Experience of measurement-based care

Regularly responding to and reviewing standardized questionnaires, while potentially familiar as a concept, was a novel experience for many physicians and patients in the study, and constituted the most significant interaction with the tool in terms of time for patients, as well as an integral part of tool use for physicians.

3.3.4. Measurement-based care: patient perspective

The majority of patients ($n = 10$) reported an increased awareness of their thoughts, feelings, and how they changed over time.

Patients generally found the questionnaires rapid and easy to complete; additionally, many reported enjoying or finding it helpful to take the time to reflect on how they were feeling (Box 1).

While 6 of 12 patients expressed, as one patient put it, that the "app covered everything [they] would have wanted to report," a number of patients did find the use of standardized questionnaires to be restrictive. As will be discussed below, this is a reasonable tension to expect when using standardized questionnaires.

3.3.5. Measurement-based care: physician perspective

One physician said that the tool "made it somewhat easier to get standardized questionnaires completed by my patients [which is an] incentive [for use]". Two other physicians agreed the questionnaires were beneficial in making patients more active in their care. Refer to *Patient agency and self-management* for specific quotes. Another said that it was "useful to monitor variations in scores". Mirroring the concerns of some patients, one physician explained that many questionnaires are not always specific to the patient. In addition, there can be tension between the results of a standardized questionnaire and clinical judgement: one doctor described that in one instance "what the patient reported to me and their score on the tool did not add up. [Their] score was higher than the symptoms the patient was presenting. [On the other hand], the numbers produced by the AI/predictive model were fine".

1 Understanding and Trust

3.3.6. Perceived level of understanding of AI

Physicians had mixed responses when asked if they understood how the AI worked. 3/7 doctors (2 psychiatrists, 1 gp) stated they understood the AI to at least a basic level, while 3/7 doctors (1 psychiatrist, 2 gp) shared they felt they did not understand the AI. Generally, physicians reported using the AI as "a supplement to clinical judgement" and, as a result, they did not place emphasis on the importance of their understanding. A key point to note here is that, even if they did not understand the technical details in full, physicians understood the *role* of the AI as an assistant to their decision making. Refer to Supplementary Material for further comments.

3.3.7. Feelings of trust in the CDSS

Physician trust in the CDSS, including the AI results, despite the mixed results regarding understanding of the AI component, was found to be generally positive. 6/7 physicians (2 psychiatrists, 4 GPs) stated that they trusted the CDSS; 1 psychiatrist with complex patients expressed hesitance in fully trusting the tool because it did not take into consideration failed treatments (as discussed above). In addition, as noted above in the section on clinical utility, the majority of physicians trusted the tool enough to use it to confirm their decisions, and 4/7 physicians reported being more open to prescribing medications they would not have considered otherwise.

Further, three doctors attributed their trust in the tool to the fact that the app mirrors the CANMAT guidelines and that the AI demonstrated "reasonable options that [they] would've suggested." 5/7 doctors reported they trusted the tool because, as one doctor put it, it "was aligned with doctor's clinical opinions and was good reinforcement from an exogenous source." Doctors felt comfortable using the app in practice because they were, as per one physician, "never surprised" by the CDSS' recommendations and the "AI explained reasoning behind its choices." A physician noted that the CDSS reported recommendations that "felt accurate and informative." Confirming the doctor's initial thinking contributed to trust in the app and its continued use.

Patient feelings of how well the CDSS represented their individual clinical situation seemed to be an important component of their overall trust in the tool. 6/12 patients reported generally trusting the tool and feeling represented by it. One patient reported that "I found that it was pretty spot on and accurate with regards to how I felt. It almost seems like it had a good idea of who I am." 6/12 patients felt reassured that the tool produced personalized results, with one patient sharing that it "felt personalized." Another two patients shared that they felt that the tool is trustworthy and a "Good tool in the hands of a good doctor. I would be a

little perturbed if a doctor based the diagnosis/medications solely on the app without going into depth with the patient." Patients felt comfortable in the tool being used by physicians but "[they] don't think the AI is capable of making a decision for the treatment of the patient.". 6/12 patients stated that they trusted their doctors to "recognize when they should trust the AI." Finally, one patient found it interesting to see their medication in the tool and to know that the choice of treatment was "backed by patient samples".

4. Discussion

In this paper, we present a mixed-methods feasibility study of novel AI-enabled CDSS. Seven physicians and 14 patients completed this feasibility study, conducted during the COVID-19 pandemic, and provided valuable insights regarding the feasibility, clinical utility and useability of the system; its impact on the physician-patient relationship; their experience of measurement-based care; and their understanding and trust in the CDSS. These different thematic categories, and relevant quantitative data, provide valuable insight into not only the integration of this particular CDSS into primary and specialty care, but into the general considerations for the integration of CDSS with AI components in mental healthcare.

4.1. Clinical feasibility, utility and useability

As noted in the quantitative results and the results of our previous paper on this study, which was focused on feasibility, (Popescu et al., 2021), both patients and physicians found the tool easy to use and feasible in a clinical setting, building on the results from our previous simulation-center based ease-of-use study (Benrimoh et al., 2021; Tanguay-Sela et al., 2022). Refer to supplementary material for further comments.

Physicians and patients agreed about the clinical utility of the application's mood tracking features to provide improved information about patient status. Physicians appreciated the clinical algorithm and the AI/predictive model; 6/7 clinicians noted that it was useful in their clinical decision-making and a majority noted its utility to confirm clinical decisions. A number of physicians also noted its ability to help them reflect on clinical decisions or consider options they otherwise would not have, which may be an important contribution given the fact that physicians tend to use the same treatments repeatedly (Frank and Zeckhauser, 2007). Given that one key objective of this tool is to help personalize treatment, it is encouraging to recognize that physicians are open to considering the information it provides and re-visiting existing patterns (Henshall et al., 2017). It is also interesting to note that updated versions of the tool may have a role to play in providing continuing medical education and that this may in turn improve the utility of the clinical decision support. This is fundamentally how the system is intended to be used: rather than physicians abdicating responsibility and simply following AI recommendations, they are meant to incorporate the predictions of the model while still considering their own expertise and collaborating with the patient.

Other comments from physicians and patients also provide valuable insight into how to continue evolving AI-enabled CDSS' to provide further clinical utility. For example, as noted above, one psychiatrist with complex patients noted that he would have preferred if the AI had taken previous treatment failure into account. While this was not possible with the current dataset and model architecture, it is something that will be taken into account during the design of future AI modules. The same psychiatrist noted that they would be more likely to use AI predictions if the predicted efficacy of treatments varied more greatly at the individual patient level. This bears further investigation, and this metric will also be used when selecting optimal models for inclusion in the CDSS in future releases. As noted above, some patients noted that they wanted to continue using the tool in a maintenance capacity, rather than with the same weekly intensity used during the study. The tool is

indeed designed to allow for a reduction in the frequency of questionnaires, the questionnaire frequency was kept constant throughout the duration of the study in order to standardize assessments across patients. Further work on features supporting patients in maintenance might also extend the useful lifespan of the tool for each patient.

Physicians noted (see supplementary materials), in a manner similar to comments made in our simulation center study (Benrimoh et al., 2020), that they felt that more practice with the tool would have improved their comfort with it and ability to use it clinically. The protocol and application training session in this study for physicians was longer than in the simulation center study (roughly 1 hour compared to roughly 40 min); however, it appears that more practice would be beneficial. As a result, in our upcoming effectiveness study clinicians will receive more detailed training and will be encouraged by local coordinators to practice using the tool in the time between initial training and patient recruitment.

4.2. Impact on physician-patient relationship

The prospect of implementing a CDSS with a novel AI component raised concerns about potential deleterious effects on the clinician-patient relationship. Indeed, one reason that this study was designed with a baseline session without the tool followed by the introduction of the tool was to provide an opportunity to note if tool introduction caused friction between patients and physicians. Encouragingly, quantitative results from the end questionnaire and STAR questionnaires demonstrate no negative effects on the relationship, and this is echoed by the qualitative results. While many clinicians and patients noted no change in the relationship- likely because, as noted by some patients above, the patients had existing positive relationships with their clinicians- there was a significant portion of patients (46 %) who noted an improvement in their relationship with their physician. As such, the tool appears to either have no negative effect on the clinician-patient relationship, or to have a positive effect for nearly half of patients. The qualitative results provide two potential mechanisms by which this positive effect might be produced. One is improved communication (Lin, 2012), with patients feeling that their physicians had a better sense of their cases and were more prepared for appointments. The other is an improved sense of patient ability to self-advocate, which corresponded to physician comments about more active patient involvement in care, having the same questionnaire results and graphs available to both parties may have helped to improve this sense of patient self-advocacy (Walker et al., 2011). Both mechanisms may have contributed to the improvements in shared decision making commented on by participants.

Additionally, the choice to introduce the tool after the first visit was made in order to demonstrate that appointment lengths did not increase after the introduction of the tool. This was in reaction to clinician comments prior to the study being designed wherein multiple clinicians noted that increased appointment length would be a significant barrier to them using the tool. However, introducing the tool after the first visit likely limited the ability of the tool to impact treatment decisions during the time of the study, given that some treatment decisions may have been made at the first visit. However, given clinician concerns, our priority in this study was to establish clinical feasibility and effect on clinician-patient relationship; definitive information about effect on treatment decisions will be established in future studies, such as our ongoing randomized clinical trial (NCT04655924).

This study was conducted during the COVID-19 pandemic; as a result, most visits were over the phone or using a teleconferencing solution. The shift to telemedicine as a result of the pandemic likely impacted the clinician and patient experience with our tool. Many visits occurred by phone, or without screen sharing functions, limiting the ability of clinicians and patients to look at the screen together. In our previous work (Benrimoh et al., 2020), we noted that standardized patients appreciated being "invited in" by clinicians to look at the tool together during decision making. As such, it is possible that the shift to

telemedicine may have reduced the positive impact of the tool. However, we note that a number of participants did express that the tool supported the transition to telemedicine, suggesting that this may be another use of the tool to be explored in future work. Most visits occurred during the pandemic, and as such data collection was mostly done remotely, reducing concerns about the equivalence of data collected in-person compared to remotely. In addition, efforts were made to ensure equivalence of data collected. For example, data on appointment lengths were recorded from door close to door open in-person, and by phone or telemedicine call recorded times remotely. It is possible that unfamiliarity with telemedicine technology may have added some time to visits, though the potential impact of this is difficult to ascertain and there were no significant variations between visit lengths at each time point.

As detailed in the Supplementary Results, many participants noted that the tool supported the transition to telemedicine.

4.3. Experience of measurement-based care

The experience of measurement-based care was overall much appreciated by both physicians and patients, despite the need to answer four questionnaires weekly in addition to a number of baseline questionnaires. As noted, 6/7 physicians felt the tool helped them get a better sense of patient situations, and as discussed above patients felt their physicians were better prepared for their appointments. Furthermore, patients noted that an important aspect of using the tool for them was the chance to self-reflect and realize how their situations have evolved over time. They noted the utility of the CDSS as effectively a neutral party they could report symptoms to that were otherwise difficult to verbalize or discuss directly in sessions. These important benefits help to explain the positive impacts of the tool noted by participants on shared decision making and communication.

While the tool seems to help patients report and reflect on their clinical symptoms and course, its reliance on standardized questionnaires, necessary for effective measurement-based care (Fortney et al., 2017), does create some expected tensions. Not all patients felt the questionnaires allowed them to fully report how they were feeling, and a number of patients suggested improvements such as daily mood tracking or a free-form diary or comments section - as one patient put it, "something a little more personalized where you can write yourself". Future versions of the tool may be able to incorporate more personalized elements, as well as to borrow tools from other methodologies, such as ecological momentary assessment, to begin personalizing patient data collection (Moskowitz and Young, 2006). This may in turn help support a wider range of therapies (i.e. patient diaries in cognitive behavioral therapy) (Nes et al., 2013).

4.4. Understanding and trust

Physician and patient understanding and trust of the app and AI were two significant factors queried in this study. Successful integration of the app and AI into standard practice is contingent on physicians', as well as the patients' perception of their physicians' understanding and trust in the tool. Physicians and patients were not given explicit instructions on how to implement the tool and AI into their care, and physicians were not told how to talk to their patients about the CDSS. As such, we were able, as in our simulation center study, to study how physicians and patients naturally learned to use the tool in their interactions and how they included it in their understanding of the treatment process.

It is interesting to observe that neither physicians nor patients endorsed fully understanding the AI and how it worked, though a number of physicians did note a basic understanding (we note that all physicians were trained on the basics of how the AI model was trained and had the opportunity to ask questions during their training session). The patient's reaction to their level of understanding was essentially to note that their main concern was their physician's ability to understand

and use the system appropriately. This underscores the importance of trust and strength in the relationship between physician and patient and, as such, makes it doubly important to demonstrate, as we have here, the lack of negative impact on the physician-patient relationship of the CDSS. Despite having only a basic understanding of the AI, the majority of physicians did endorse trusting the CDSS as a whole. Some clinicians noted that this was because the majority of information presented in the application (with the exception of predicted remission probabilities) came from established treatment guidelines, underscoring the value of the design of this CDSS, which sought to use AI to build upon the gold-standard approach. While they may not have understood all the *technical* details of the AI, clinicians *did* understand the way in which the AI was meant to be used- as an aid to clinical decision, as one more piece of information they could consider (Jeffries et al., 2021). Most physicians had only a few patients in this study. As such, one might argue that what was observed in this study is the 'exploratory' phase of physicians integrating this new technology - with physicians willing and able to use it, but still learning how it works and how best to leverage it in their practices. Given that technology remains quite novel to physicians, with a systematic review reporting that 70.2 % of physicians had never used applications in clinical practice before (Kerst et al., 2020), it is important to provide continuing education and training for healthcare providers (Briganti and Le Moine, 2020). Our upcoming effectiveness trial, which asks physicians to enroll double the number of patients as in this study, may therefore provide more information about how physicians use this technology with more practice and practical experience with the system.

5. Limitations

This study has a number of important limitations. One key limitation was the small sample size, which may limit the generalizability of results. In addition, a key limitation was that resource considerations limited us to the use of written notes rather than recordings. Written notes rely on the capacity of each interviewer to correctly record the comments and as such might result in reduced accuracy compared to recordings. In addition, the need to write while listening to the participant may lead to some statements being missed, resulting in a loss of richness in the data. We attempted to mitigate this by using a semi-structured interview, which allowed for the same questions to be asked in the same way to all participants, while allowing the RA to ask clarifying or exploratory questions when they deemed it relevant. In addition, RAs were given clear instructions to record all statements verbatim. The study was designed with a focus squarely on feasibility, with the tool being introduced only after the first visit with the patient. Indeed even patients noted this, with one patient mentioning how their medications had been changed prior to the introduction of the tool and that as such the CDSS likely did not have an impact at the level of treatment choice. We note that this was not intentional, as this study was not meant to test tool effectiveness but rather to determine if its introduction was feasible or if it had any deleterious effects on the clinician-patient relationship. This limitation is further discussed in Popescu et al., 2021. Finally, as noted above this study was adjusted in order to conduct it entirely virtually, due to the COVID-19 pandemic. One negative aspect of telemedicine was reduced opportunity for physicians to look at their screens together with patients, as many of the telemedicine follow-ups were conducted over the phone. During the simulation center study of this tool (Benrimoh et al., 2020), looking at screens together was an important part of the patient-physician interaction when using the tool. As such, this study may have benefitted from more in-person visits; on the other hand, the significantly positive experience of the tool supplementing telemedicine demonstrates both the versatility of the CDSS and the manner in which patients and clinicians can adapt its use to meet present circumstances.

6. Conclusion

In conclusion, this paper has described a mixed-methods approach to the study of a novel AI-enabled CDSS for the treatment of major depression. Key themes regarding clinical utility, ease of use, feasibility, the impact on the physician-patient relationship, the experience of measurement-based care, and physician and patient understanding of and trust in the tool were discussed. Importantly, clinicians saw the tool as being a trustworthy support to help them explore more treatment options, and patients found the experience of measurement-based care to be one that helped improve self-reflection and support self-advocacy. A number of key learnings, such as the need for increased training time and the importance of administrative support, will help support future effectiveness trials of this - and potentially, of other - clinical decision support systems in mental healthcare and beyond. Next steps also include upgrades to the AI model, such as an AI that takes individual failed treatment history and patient trajectories into account.

CRediT authorship contribution statement

S.Q. lead the formal analysis of the qualitative portion of the data and contributed to manuscript writing– original draft, and writing– review and editing. D.B. had a role in the conceptualization of the study, data curation, formal analysis, funding acquistion, investigation, methodology, project administration, creation of the AI model/software used, supervision, validation, writing– original draft, and writing– review and editing. C.P. had a role in study administration, data curation, formal analysis, and coordinate the study, participated in data collection and analysis, and writing– original draft, and writing– review and editing. M.T.S. helped with study administration and data curation. E.L, J.B, and B.D participated in data curation and formal analysis. M.S., K. W., T. P, E.S., E.L, and MJ.F. participated in formal analysis and contributed to manuscript writing– original draft, and writing– review and editing. S.I., K.P., contributed to study conceptualization and to manuscript writing– review and editing. C.A., R.F., J.M., participated in the creation of the software/AI model used for this study and in manuscript writing– review and editing. G.G., and D.S. participated in formal analysis, manuscript writing– original draft, and writing– review and editing. W.S., S.R., and K.L. were part of the study data safety monitoring board. K.L., W.S., S.R., J.K., K.H., S.P., R.M.S, and M.F. provided guidance on the study implementation and on the manuscript development as well as manuscript writing– review and editing. H.M. provided supervision, participated in the conceptualization of the study, and participated in manuscript writing – review and editing.

Role of the funding source

Funding was obtained between initial IRB approval (July 29th, 2019) and the start of the study but was disclosed to REB as amendments were approved prior to study start.

Declaration of Competing Interest

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

D.B., C.A., R.F., S.I., and K.P. are shareholders and either employees, directors, or founders of Aifred Health. J.M. is employed by Aifred Health. M.T.S. is employed by Aifred Health and is an options holder. S. Q., G.G., D.S., M.S., K.W., T.P., E.S., E.L., MJ. F., J.B., B.D., and C.P., have been or are employed or financially compensated by Aifred Health. S.P., K.H., J.K. are members of Aifred Health's scientific advisory board and have received payments or options. W.S. is a member of the data safety monitoring board. H.M. has received honoraria, sponsorship or grants for participation in speaker bureaus, consultation, advisory board meetings and clinical research from Acadia, Amgen, HLS Therapeutics, Janssen-Ortho, Mylan, Otsuka-Lundbeck, Perdue, Pfizer, Shire and

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jadr.2023.100677.

References

American Psychiatric Association, 2013. DSM-5 diagnostic classification. Diagnostic and Statistical Manual of Mental Disorders, 10th ed.

Benrimoh, D., Fratila, R., Israel, S., Perlman, K., Mirchi, N., Desai, S., Rosenfeld, A., Knappe, S., Behrmann, J., Rollins, C., You, R.P., 2018. Aifred health, a deep learning powered clinical decision support system for mental health. InThe NIPS'17 Competition: Building Intelligent Systems. Springer, Cham, pp. 251–287.

Benrimoh, D., Tanguay-Sela, M., Perlman, K., Israel, S., Mehltretter, J., Armstrong, C., Fratila, R., Parikh, S.V., Karp, J.F., Heller, K., Vahia, I.V., 2020. Using a simulation centre to evaluate the effect of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. medRxiv. Jan 1.

Benrimoh, D., Tanguay-Sela, M., Perlman, K., Israel, S., Mehltretter, J., Armstrong, C., et al., 2021. 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), e22.

Bohn, M.J., Babor, T.F., Kranzler, H.R., 1995. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J. Stud. Alcohol. 56 (4), 423–432. https://doi.org/10.15288/jsa.1995.56.423.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Brigant, G., Le Moine, O., 2020. Artificial intelligence in medicine: today and tomorrow. Front. Med. (Lausanne) 7, 27. https://doi.org/10.3389/fmed.2020.00027.

Canadian Institutes of Health Research, 2005. Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

Felitti, V.J., Anda, R.F., Nordenberg, D., Williamson, D.F., Spitz, A.M., Edwards, V., Marks, J.S., 1998. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the Adverse Childhood Experiences (ACE) Study. Am. J. Prev. Med. 14 (4), 245–258. May 1.

Fortney, J.C., Unützer, J., Wrenn, G., Pyne, J.M., Smith, G.R., Schoenbaum, M., Harbin, H.T., 2017. A tipping point for measurement-based care. Psychiatr. Serv. 68 (2), 179–188. https://doi.org/10.1176/appi.ps.201500439.

Frank, R.G., Zeckhauser, R.J., 2007. Custom-made versus ready-to-wear treatments: behavioral propensities in physicians' choices. J. Health Econ. 26 (6), 1101–1127. https://doi.org/10.1016/j.jhealeco.2007.08.002.

Goldberg, S.B., Buck, B., Raphaely, S., Fortney, J.C., 2018. Measuring psychiatric symptoms remotely: a systematic review of remote measurement-based care. Curr. Psychiatry Rep. 20 (10), 81. https://doi.org/10.1007/s11920-018-0958-z.

Henshall, C., Marzano, L., Smith, K., Attenburrow, M.J., Puntis, S., Zlodre, J., Kelly, K., Broome, M.R., Shaw, S., Barrera, A., Molodynski, A., 2017. A web-based clinical decision tool to support treatment decision-making in psychiatry: a pilot focus group study with clinicians, patients and carers. BMC Psychiatry 17 (1), 1-0.

Jeffries, M., Salema, N.E., Laing, L., Shamsuddin, A., Sheikh, A., Avery, A., Chuter, A., Waring, J., Keers, R.N., 2021. The implementation, use and sustainability of a clinical decision support system for medication optimisation in primary care: a Qualitative evaluation. PLoS ONE 16 (5), e0250946.

Katz, A.J., Dusetzina, S.B., Farley, J.F., Ellis, A.R., Gaynes, B.N., Castillo, W.C., Stürmer, T., Hansen, R.A., 2012. Distressing adverse events after antidepressant switch in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial: influence of adverse events during initial treatment with citalopram on development of subsequent adverse events with an alternative antidepressant. Pharmacotherapy 32, 234–243.

Kennedy, S.H., Lam, R.W., McIntyre, R.S., Tourjman, S.V., Bhat, V., Blier, P., Hasnain, M., Jollant, F., Levitt, A.J., MacQueen, G.M., McInerney, S.J., McIntosh, D., Milev, R.V., Müller, D.J., Parikh, S.V., Pearson, N.L., Ravindran, A.V., Uher, R., CANMAT Depression Work Group, 2016. 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. Psychiat. Revue canadienne de psychiatrie. 61 (9), 540–560. https://doi.org/10.1177/0706743716659417.

Kerst, A., Zielasek, J., Gaebel, W., 2020. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur. Arch. Psychiatry Clin. Neurosci. 270 (2), 139–152. https://doi.org/10.1007/s00406-018-0974-3.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lin, H.C., 2012. Real-time clinical decision support system. Med. Inform. 111–136.

McGuire-Snieckus, R., McCabe, R., Catty, J., Hansson, L., Priebe, S., 2007. A new scale to assess the therapeutic relationship in community mental health care. STAR. PsycholMed. 37 (1), 85–95. https://doi.org/10.1017/s0033291706009299.

Mehltretter, J., Fratila, R., Benrimoh, D., Kapelner, A., Perlman, K., Snook, E., Israel, S., Armstrong, C., Miresco, M., Turecki, G., 2020. Differential treatment benefit prediction for treatment selection in depression: a deep learning analysis of STAR*D and CO-MED data. Computat. Psychiatry 4, 61–75. https://doi.org/10.1162/cpsy_a_00029.

Merlhiot, G., 2014. Adaptation and validation of the standardized assessment of personality – abbreviated scale as a self-administered screening test (SA-SAPAS). J. Psychol. Psychother. 04 (06) https://doi.org/10.4172/2161-0487.1000164.

Moskowitz, D.S., Young, S.N., 2006. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. J. Psychiatry Neurosci. 31 (1), 13–20.

Nes, A.A., Eide, H., Kristjánsdóttir, Ó.B., van Dulmen, S., 2013. Web-based, self-management enhancing interventions with e-diaries and personalized feedback for persons with chronic illness: a tale of three studies. Pat. Educ. Couns. 93 (3), 451–458. https://doi.org/10.1016/j.pec.2013.01.022.

Patel, S.R., Bakken, S., Ruland, C., 2008. Recent advances in shared decision making for mental health. Curr. Opin. Psychiatry 21 (6), 606–612. https://doi.org/10.1097/YCO.0b013e32830eb6b4.

Popescu, C., Golden, G., Benrimoh, D., et al., 2021. 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 Format. Res. 5 (10), e31862. https://doi.org/10.2196/31862.

Rush, A.J., Carmody, T., Reimitz, P.E., 2000. The inventory of depressive symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int. J. Method Psychiatr. Res. 9 (2), 45–59.

Scott, K., Lewis, C.C., 2015. Using measurement-based care to enhance any treatment. Cogn. Behav. Pract. 22 (1), 49–59. https://doi.org/10.1016/j.cbpra.2014.01.010.

Sim, I., Gorman, P., Greenes, R.A., Haynes, R.B., Kaplan, B., Lehmann, H., Tang, P.C., 2001. Clinical decision support systems for the practice of evidence-based medicine. J. Am. Med. Inform. Assoc. 8 (6), 527–534. https://doi.org/10.1136/jamia.2001.0080527.

Skinner, H.A., 1982. The drug abuse screening test. Addict. Behav. 7 (4), 363–371. https://doi.org/10.1016/0306-4603(82)90005-3.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092. https://doi.org/10.1001/archinte.166.10.1092.

Tanguay-Sela, M., Benrimoh, D., Popescu, C., Perez, T., Rollins, C., Snook, E., Lundrigan, E., Armstrong, C., Perlman, K., Fratila, R., Mehltretter, J., Israel, S., Champagne, M., Williams, J., Simard, J, Parikh, S.V., Karp, J.F., Heller, K., Linnaranta, O., Cardona, L.G., Turecki, G., Margolese, H.C., 2022. Evaluating the perceived utility of an artificial intelligence-powered clinical decision support system for depression treatment using a simulation center. Psychiatry Res. 308, 114336 https://doi.org/10.1016/j.psychres.2021.114336.

Tran, B.X., McIntyre, R.S., Latkin, C.A., Phan, H.T., Vu, G.T., Nguyen, H., Gwee, K.K., Ho, C., Ho, R., 2019. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int. J. Environ. Res. Public Health 16 (12), 2150. https://doi.org/10.3390/ijerph16122150.

Üstün, T.B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., et al., 2010. Developing the World Health Organization disability assessment schedule 2.0. Bull World Health Organ. 88 (11), 815–823.

Walker, J., Leveille, S.G., Ngo, L., Vodicka, E., Darer, J.D., Dhanireddy, S., Elmore, J.G., Feldman, H.J., Lichtenfeld, M.J., Oster, N., Ralston, J.D., Ross, S.E., Delbanco, T., 2011. Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. Ann. Intern. Med. 155 (12), 811–819. https://doi.org/10.7326/0003-4819-155-12-201112200-00003.

Wisniewski, S.R., Rush, A.J., Balasubramani, G.K., Trivedi, M.H., Nierenberg, A.A., 2006. Self-rated global measure of the frequency, intensity, and burden of side effects. J. Psychiatr. Pract. 12 (2), 71–79. https://doi.org/10.1097/00131746-200603000-00002.

Yeung, A., Feldman, G., Pedrelli, P., Hails, K., Fava, M., Reyes, T., Mundt, J.C., 2012. The quick inventory of depressive symptomatology, clinician rated and self-report: a psychometric assessment in Chinese Americans with major depressive disorder. J. Nerv. Ment. Dis. 200 (8), 712.