Scalable precision psychiatry with an objective measure of psychological stress: Prospective real-world study

Helena Wang, Norman Farb, Bechara Saab

Submitted to: Journal of Medical Internet Research
on: January 15, 2024

Table of Contents

Original Manuscript **5**
Supplementary Files **36**
Figures **37**
Figure 1 .... **38**
Figure 4 .... **43**
TOC/Feature image for homepages **44**
TOC/Feature image for homepage 0 .... **45**

Scalable precision psychiatry with an objective measure of psychological stress: Prospective real-world study

Helena Wang[1] MSc; Norman Farb[2] PhD; Bechara Saab[1] PhD

[1] Mobio Interactive Pte Ltd Singapore SG
[2] University of Toronto Toronto CA

Corresponding Author:
Bechara Saab PhD

Mobio Interactive Pte Ltd
726A Geylang Road
Singapore
SG

Abstract

Background: Precision psychiatry is data-driven personalisation of mental healthcare. One major aim is to identify the specific forms of therapy that will be most clinically meaningful for each patient. Before major progress towards this aim is possible however, tools that objectively assess the mental wellbeing of a patient must be validated in this context.

Objective: To compare the fidelity of a therapy recommendation algorithm when trained with an objective quantification of psychological stress versus subjective ecological momentary assessments of stress and mood.

Methods: From 2,786 unique individuals engaging between March 2015 and December 2022 in English language psychotherapy sessions and providing pre- and post-session self-report and facial biometric data via the AmDTx mental health platform (Mobio Interactive Pte Ltd, Singapore), analysis was conducted on 67 "super users" that completed at least 28 sessions with all pre- and post-session measures. AmDTx is a clinically validated mental health platform that provides patients with audio recordings supporting mental wellbeing (asynchronous and on-demand psychotherapy). AmDTx also contains easy to use tools that rapidly assess mental wellbeing, including an objective measure of psychological stress derived from AI analysis of facial biomarkers (Objective Stress Level; ?OSL), and ecological momentary assessments (EMAs). Two commonly used EMAs within AmDTx are self-reported stress (?SRS) and self-reported mood (?SRM). These three data sources were used to independently train an algorithm designed to predict what future therapy sessions would prove most efficacious for each individual. Algorithm predictions were compared against the efficacy of the individual's self-selected sessions.

Results: EMAs demonstrated significant convergence with each other ($r=0.53$, $P<.01$), and divergence from the objective measure of psychological stress ($r<0.03$). The recommendation algorithm selected increasingly efficacious therapy sessions as a function of training data, when trained with the pre- and post-session ?OSL ($F(1, 16)=5.37$, $P=.034$) or ?SRM data ($F(1, 16) = 3.69$, $P=.048$). Using the first 15 therapy sessions, the algorithm trained using ?OSL data predicted future sessions with a statistically higher efficacy relative to self-selected sessions ($P<.01$). Training the algorithm with self-reported data showed potential trends that did not reach significance (?SRS: $P=.092$; ?SRM: $P=.12$). Self-selections were over-represented amongst the algorithmically recommended sessions, an effect most pronounced for ?OSL ($F(1, 14)=30.94$, $P<.001$).

Conclusions: A rapid, scalable, and objective measure of psychological stress, when combined with a robust recommendation algorithm, may improve clinical predictiveness within precision psychiatry and thereby support the capacity for healthcare professionals to meet global demand. Clinical Trial: clinicaltrials.gov

(JMIR Preprints 15/01/2024:56086)
DOI: https://doi.org/10.2196/preprints.56086

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Scalable precision psychiatry with an objective measure of psychological stress: Prospective real-world study

Helena Wang [1], Norman A. S. Farb [2,3], Bechara Saab [1]*

[1] Mobio Interactive Pte Ltd, 726A Geylang Road, Singapore 389637
[2] Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada, L5L 1C6
[3] Department of Psychological Clinical Sciences, University of Toronto Scarborough, Scarborough, Ontario, Canada, M1C 1A4

*Corresponding Author
Bechara Saab, PhD
Mobio Interactive Pte Ltd
726A Geylang Road
Singapore 389637
bechara@mobiointeractive.com

Abstract

Background: Before meaningful progress towards precision psychiatry is possible, objective (unbiased) assessment of patient mental wellbeing must be validated and adopted broadly in this context.

Objective: To compare the fidelity of a precision psychiatry therapy recommendation algorithm when trained with an objective quantification of psychological stress *versus* subjective ecological momentary assessments (EMAs) of stress and mood.

Methods: From 2,786 unique individuals engaging between March 2015 and December 2022 in English language psychotherapy sessions and providing pre- and post-session self-report and facial biometric data via a mobile health platform (Mobio Interactive Pte Ltd, Singapore), analysis was conducted on 67 "super users" that completed a minimum of 28 sessions with all pre- and post-session measures. The platform employed has been previously shown to reduce psychiatric symptom severity and improves overall mental wellbeing via asynchronous and on-demand psychotherapy. Psychotherapy recordings ("sessions") within the platform span mindfulness, meditation, cognitive behavioral therapy, client-centered therapy, music therapy, self-hypnosis. The platform also has the unusual ability to rapidly assess mental wellbeing without bias, via an easy-to-use objective measure of psychological stress derived from AI analysis of facial biomarkers (Objective Stress Level; OSL). In tandem with the objective measure, EMAs obtain self-reported values of stress (SRS) and mood (SRM). ΔOSL, ΔSRS, and ΔSRM (with delta referring to the pre-session subtracted from the post-session measurement) were used to independently train a therapy recommendation algorithm designed to predict what future sessions would prove most efficacious for each individual. Algorithm predictions were compared against the efficacy of the individual's self-selected sessions.

Results: The objective measure of psychological stress provided a differentiated delta for the measurement of therapeutic efficacy compared to the two EMA deltas, as shown by clear divergence in ΔOSL vs ΔSRS or ΔSRM ($r<.03$), while the EMA deltas showed significant convergence with each other ($r=.53$, $P<.01$). The recommendation algorithm selected increasingly efficacious therapy sessions as a function of training data, when trained with either ΔOSL ($F(1, 16)=5.37$, $P=.034$), or ΔSRM data ($F(1, 16) = 3.69$, $P=.048$). However, the sequential improvement in prediction efficacy only surpassed the efficacy of self-selected therapy when the algorithm was trained using objective data ($P<.01$; using 15 training sessions). Training the algorithm with EMA data showed potential trends that did not reach significance (ΔSRS: P=.092; ΔSRM: P=.12). A final, insightful finding is that self-selected therapy sessions were over-represented amongst the algorithmically recommended sessions, an effect most pronounced when the algorithm was trained with ΔOSL data ($F(1, 14)=30.94$, P<.001).

Conclusions: These prospective data demonstrate that a rapid, scalable, and objective measure of psychological stress, in combination with a robust recommendation algorithm, can autonomously identify clinically meaningful therapy for individuals. More broadly, this work illustrates the potential for objective data on mental wellbeing to improve precision psychiatry, and the capacity for mental healthcare professionals to match global demand.

Trial Registration: https://clinicaltrials.gov/ct2/show/NCT06265909

Key Words

Precision Psychiatry; Digital Biomarkers; Personalization; Mental Health; Digital Therapeutics; Recommender System; Psychological Stress; Therapy Algorithm

Introduction

Psychiatry remains unique within medicine as the only major field that does not employ objective data in standard practice.

Precision psychiatry aims to change this, pulling psychiatry into the realm of modern medicine with treatment options that match the unique profile of each patient [1]. Where employed, these personalized treatment plans are predominantly informed by neuroimaging, genetic biomarkers, and medical history. However, despite great promise and considerable funding, precision psychiatry has yet to deliver major improvements for the patient [2,3]. In this study, we explore if readily obtainable objective data on the patient's moment-to-moment mental wellbeing can be used to inform a precision psychiatry therapy recommendation algorithm.

At present, mental wellbeing and psychiatric symptom severity are predominantly assessed through retrospective self-reports in the form of psychological scales. While these scales are often rigorously interrogated for face, content, and divergence reliability and validity, it is well accepted that they suffer from profound limitations when employed in the real world. These multifaceted limitations fall into three general classes that serve as barriers, obstacles, and misdirection preventing healthcare professionals from accessing the true mental experience of their patients (Figure 1A).

First, scales have a "temporal" barrier, typically requiring about 5 to 30 min to complete, and designed to retrospectively access a relatively wide window of time (e.g., the previous two weeks). Psychological scales, therefore, cannot be administered in rapid succession or assess real time changes in a patient's mental wellbeing [4]. Second, scales impart "ethnological" obstacles since natural cultural and personality divergence means that any given psychological scale will only be maximally informative for a relatively narrow portion of the global population [5]. Third, and both most obvious and most troubling, scales can be misdirected by "bias". Subjective by design, scales contain a plethora of documented biases related to i) recency, ii) recall, iii) expectation, iv) confirmation, and v) response [6,7]. Response biases, in particular with respect to gender [8] and cultural background [9,10], markedly influence the tendency to report negative affect.

Temporal barriers of psychological scales can be addressed by incorporating ecological momentary assessments (EMAs), a practice of daily or hourly self-reporting often completed by the patient via a common mobile device [11]. However, subjective by nature, EMAs do little to address the other two limitations of subjective measures. To address ethnological and bias limitations, objective data are essential. Already precision psychiatry attempts to incorporate many objective measures, in particular brain imaging and (epi-)genotyping [3]. However, these data, like medical records, are also susceptible to temporal limitations. Moreover, the objective data currently used, and proposed for use, in precision psychiatry [12,13], even if collected through accessible tools such as wearables, may be too far removed from the phenotype of clinical interest (mental wellbeing) to optimally inform precision psychiatry models (Figure 1B). We argue here that a rapid pan-ethnological and objective measure of mental wellbeing is needed to meaningfully move towards precision psychiatry that is viable at the scale required to match global demand.

Understanding the profound medical opportunity that would emerge from establishing measures of mental health that are not subject to the temporal, ethnological, and bias limitations, we developed a novel objective measure of psychological stress using deep neural network (DNN) processing of selfie video biomarkers [14]. We chose stress for two reasons. Frist, stress is a physiological process tightly linked to overall wellbeing [15], and directly causes or negatively impacts a broad proportion and large variety of medical conditions [16]. Second, the autonomic nervous system underlying stress influences heart rate variability (HRV), which can be directly measured at scale with remote photoplethysmography (rPPG) [17,18]. Indeed, due to its relationship with stress, HRV has recently gained traction in medical practice [19,20]. Our psychological stress DNN was trained with data from thousands of patients in over 150 countries, from whom heart rate and the power of HRV within two frequency domains (“HRV-high” and “HRV-low") [14] were captured in real time via facial rPPG [17]. In a previous analysis, this DNN achieved 86% accuracy and a mean squared error of 0.01, substantially outperforming the logistic regression conventionally employed to estimate stress from HRV [14].

The purpose of our current study was to probe the potential utility of the DNN psychological stress data for training a precision psychiatry recommendation algorithm. The ultimate aim is to improve clinical efficacy by rapidly matching patients to their ideal therapy regimen. To this end, we made use of a commercialized mobile health platform (Mobio Interactive Pte Ltd, Singapore). The platform has been used in a variety of clinical contexts [21-33] and is one of the few software applications of its kind to outperform placebo in a randomized controlled trial [21]. The platform gives patients access to over 1,000 distinct audio files delivered by over 60 professionals and across six major languages as of January 2024. These audio files are typically 5 to 45 min sessions that serve to support relaxation, enhance stress resilience, and otherwise stimulate neural plasticity underlying psychiatric symptom reduction. Since the mobile platform is installed on a patient’s personal smartphone or tablet, patients are free to engage with therapy at the time and place most convenient and/or effective for them personally. Before and after these on-demand asynchronous sessions, patients are invited to perform a 30 s selfie scan (capturing their psychological stress) and complete two EMAs, previously demonstrated to correlate with high significance to the principal components of relevant psychological scales [21]. The platform therefore provides an attractive means to investigate the fidelity of therapy recommendation algorithms.

Use of recommendation systems within mobile apps is of course not new e.g., [34-36]. However, to our knowledge there have been no attempts to train a recommendation algorithm with an objective measure of mental wellbeing. In a recent review of 73 studies on recommendation systems in health apps, approximately half (47%) did not attempt to validate the recommendations [37]. Those that did primarily employed self-reported measures of mood, sleep, anxiety, depression, and/or energy; and compared users who received personalized content recommendations to passive control groups that either received no content or received content at random. Seldom have objective data been used to train or evaluate recommendation algorithms, and in these cases the results are thus far underwhelming. For example, tailored antismoking messages to patients with substance use disorder led to higher user ratings (subjective data) when compared to a conventional messaging system, but the tailored recommendations failed to impact smoking rates (objective data) [38].

Here we evaluate the fidelity of a precision psychiatry recommendation algorithm when trained with objectively quantified stress, in comparison to training with two validated EMAs of stress and mood. In all cases, the efficacy of the recommended therapy was examined relative to the efficacy of therapy content that was self-selected by patients. This prospective study design provides the controlled methodology required to model the impact of precision psychiatry against standard care.

Methods

Study design, Patient and Public Involvement

The prospective, observational real-world study accessed data from a mobile health platform (Mobio Interactive Pte Ltd, Singapore) equipped with computer vision and AI to objectively quantify psychological stress [14,17] and benchmarked EMAs to subjectively measure stress, valence, and arousal [21]. Asynchronous and on-demand psychotherapy is available as audio files and has been clinically validated across the mental illness severity spectrum [21-24,26,27,30-33]. Users consented to the use of their anonymous data for research purposes via the platform Terms of Use and were not involved in a formal process to inform the design of this research. Patient identity remained anonymous throughout the study. IRB exception for consent was granted by Avdarra (Ontario, Canada).

Data collection, Setting, Participants, Bias, Study size

Data were collected between March 2015 and December 2022 on 36,160 unique users in a manner compliant with the Health Insurance Portability and Accountability Act (HIPAA), Personal Health Information Protection Act (PHIPA), and General Data Protection Regulation (GDPR). Of these, 2,786 unique individuals engaged in biometric and self-report data collection before and after engaging with asynchronous psychotherapy. No audio or video were collected, transferred, or stored.

To protect the real-world applicability of the results, users were not given any special instruction or information about the nature or possibility of the current analyses. As a consequence, user data varied greatly in terms of engagement and app-use characteristics. To create a single, unified, and consistent dataset that could be leveraged across all intended analyses, data were filtered to only include English-language psychotherapy sessions that contained the required session payloads for algorithm inclusion (see below), and only when the objective and two subjective measures were all completed both before and after each session. A power analysis indicated a minimum of 66 unique users are required to achieve 80% power for detecting a medium effect size via the intended analyses. This limit afforded over 28 sessions with all pre- and post-session measures from 67 unique users.

Materials and variables

Objective stress level (ΔOSL). Objective stress was obtained via a 30-second "selfie" video captured with the front-facing camera of a mobile device (smartphone or tablet). The computer vision data extracted from the videos in real time were then passed through a deep neural network (DNN) to compute the objective stress level (ΔOSL) at that moment in time, as previously described [14]. ΔOSL is represented with a value between 0 to 1, with greater values representing more stress.

Subjective, self-reported stress (ΔSRS). Subjective stress was quantified via an animated digital "slider". Users reported their current level of stress either by dragging a marker on the slider to a position of their choosing between "none" (0) and "extreme" (10), or by tapping on one of four faces positioned above the slider, with each face visually depicting stress levels at the mid-points of four quadrants (i.e., values of 1.25, 3.75, 6.25, 8.75). Users were instructed to input the stress that represents how they feel "right here, right now".

Subjective, self-reported mood (ΔSRM). Subjective mood was quantified via a "mood board", which asks users to select from 32 different words representing various emotions (e.g., "delighted", "content", "gloomy", "tense"). The mood board consists of two axes, one spanning from "unpleasant" to "pleasant" and the other from "mild" to "intense". Each quadrant contains 8 mood words. Users were instructed to tap on the words that represent how they feel "right here, right now". Unlike ΔSRS, where users directly report a numerical value representing their stress level, subjective mood is quantified indirectly from the selected words. During data analysis, each selected word was assigned a score between -2 and +2, depending on its connotations with respect to valence and arousal/saliency. For example, the mood words "happy", "relaxed", "depressed", and "nervous" were assigned scores of 2, 1, -1, and -2, respectively. An overall score of mood was then computed from the sum of the selected mood words, and converting this to a value between 0 to 1, where a greater score indicates better mood. ΔSRM is therefore an implicit measure in so far as users are not informed how selecting words informs the overall score of their mood.

Recommendation algorithm

The recommendation algorithm was developed internally at Mobio Interactive in advance of conducting this study and was not altered during analysis. The algorithm uses a non-collaborative user-item interaction design, relying exclusively on pre/post session data from a consistent user. Prior knowledge of user characteristics (e.g., gender, age, ethnicity, medical diagnosis) is not necessary. The content selections and responses of other users were not considered, and thus did not influence the algorithm's recommendations for a given user. This algorithm design thus leverages neither collaborative filtering nor content-based filtering, instead treating each individual as an individual without prior assumptions about how therapy may influence them based on demographics, diagnoses, or the response patterns observed at large. A design of this nature was considered necessary due to the global footprint of the dataset and an understanding that how patients respond to treatment is highly idiosyncratic and unpredictable. Thus, the only assumption used by the algorithm is that the forms of therapy that are most effective for a given patient/user at one point in time, will reliably predict the forms of therapy that will be most effective for the same patient/user at a later point in time.

Three types of pre/post session data were available to provide an evaluation of historical efficacy: ΔOSL, ΔSRS, and ΔSRM (as explained above). The algorithm was trained independently on each of these three measures, first by calculating the post-pre delta:

$$\Delta WM = WMpost - WMpre \tag{1}$$

where WM is the wellbeing measure (i.e., ΔOSL, ΔSRS, or ΔSRM). For measures of stress, the more negative the ΔWM the more efficacious the session is considered for that user (and vice versa for mood). For each user, all sessions were then ranked according to efficacy.

Sessions that are intended to be directly therapeutic i.e., are not purely (psycho)educational, are each paired with i) three out of a total of 36 possible “mood words” and ii) three out of a total of 24 possible “intent words”. The pairing of sessions to mood and intent words was conducted according to a consensus amongst the content creation team at Mobio Interactive, and was later verified by the session’s guide (the voice within each audio recording), including clinical psychologists, psychiatrists, mindfulness teachers, and/or other relevant professionals involved in content creation for the platform. Mood and intent words paired to a given session were next assigned a “raw word score”, R equal to the ΔWM value from that session.

Next, the algorithm then allowed for potentially important factors of influence to scale R:

$$S = R \times A \times B \times C \ldots \tag{2}$$

where S is the scaled word score, and A, B, and C are scalers. At the time of analysis, one scaler was used, A, to control for the potential influence of the session guide, and according to the formula:

$$A = \left(\frac{\underline{G_x} - \underline{G_{min}}}{\underline{G_{max}} - \underline{G_{min}}} \right) \times W + 1 \tag{3}$$

Where $\underline{G_x}$ is the mean ΔWM of the session guide for the user, $\underline{G_{min}}$ is the mean ΔWM of sessions delivered by the guide that is least efficacious for the user, $\underline{G_{max}}$ is the mean ΔWM of sessions delivered by the guide that is the most efficacious for the user, and W is the “guide weight”, an arbitrary constant that controls the degree to which scaler A influences S. For simplicity, this analysis used $W = 1$. Formula 3 increases the influence of mood and intent words delivered by session guides that are generally more efficacious for a given user.

Next, for each mood and intent word, S for each session was summated to produce the "total word score", T:

$$T = \sum S \qquad \{3\}$$

Finally, the three mood words and three intent words for each user with the greatest T were input into a proprietary algorithm that ranks sessions within the entire possible library according to their association with that specific combination of six words. The top three sessions in this ranking were considered the algorithmically selected (AS) sessions for the patient.

Evaluation of the algorithm. The algorithm, designed to use each user's corpus of session data to rank the likelihood of a given session for its potential to benefit the same user, theoretically should increase its predictiveness as the user completes more sessions. To explore the effect of increasing training set size, therefore, we trained the algorithm multiple times for each user, starting with one session and sequentially increasing the number of training sessions by one. At each stage, algorithm predictiveness was evaluated using all remaining session data from the same user. For example, if a user completed a total of 75 sessions and the first 15 were used to train the algorithm, then sessions 16 to 75 were used for testing. If a user completed a total of only 35 sessions and the first 10 were used to train the algorithm, then sessions 11 to 35 were used for testing. Increasing training set size thus had an inverse impact on testing set size. The sequentially decreasing size of the testing set gave rise to statistical limitations at the tail end of this process. However, this analysis design was still preferrable to alternatives since it maintained cross-comparison consistency, maximized the total size of the dataset available for analysis, and best reflects the real-world nature of the source data.

In all cases, the data type (ΔOSL *vs.* ΔSRS *vs.* ΔSRM) used to train the algorithm was also used to evaluate the predictiveness of the algorithm. To examine if AS sessions were associated with a better outcome for the user, they were compared against "user selected" (US) sessions using a within-subject ANOVA. The choice to include all user selected sessions (including those that could have been simultaneously selected by the algorithm) was made to ensure the highest bar for comparison. The comparison design may underestimate the difference between AS and US sessions, but nonetheless best reflects the real-world nature of the source data and the real-world application of these types of healthcare solutions. Comparison against self-selection was thus, in our view, the only meaningful choice. To compare results across various wellbeing measures, data were z-score normalized.

Cross-validation. ANOVA comparisons were 10-fold cross-validated using a bootstrapping approach by randomly sampling 80% of the users (54 users) and repeating the same training and testing on this reduced sample.

Statistical analysis software

All statistical analyses were conducted using R (version 4.0.2).

Results

Participants and descriptive data

Of the 67 patients eligible for analysis, 43 (64%) identified as female, 19 (28%) identified as male, and 5 (7%) selected “other” under gender. Self-reported age ranged from 18 to 66, with the largest age group being 25-34 years old (n=21, 31%) and the smallest age group being 18-24 years old (n=7, 10%). Geolocation data indicated predominant use in Australia (n=27, 40%), New Zealand (n=12, 18%), and Canada (n=11, 16%). The mean number of sessions completed by each user (with or without pre- and post-session measures) was 154.9 (sd=228.1), while the median number and range of sessions completed were 76 and 40-1328, respectively. English-language sessions with all pre- and post-session measures completed by the participants had a mean of 42.7 (sd=16.8), median of 37, and range of 28-108.

Asynchronous on-demand therapy is beneficial for patients

Inspection of the three measures intended to train the recommendation algorithm suggested that across all sessions, there was a mean reduction in stress and improvement in mood, since in all cases the confidence intervals did not include zero (ΔOSL=-0.003; 95% CI [-0.00091,-0.026]; ΔSRS=-0.09; 95% CI [-0.0075,-0.12]; ΔSRM=0.13; 95% CI [0.012,0.24]]). These real-world data corroborate controlled clinical investigations [21-24,26,27,30-32].

Divergence between objective and subjective data

Pearson correlations between the three measures intended to train the recommendation algorithm suggested independence between the objective and the two subjective measure deltas. ΔSRS and ΔSRM correlated significantly with each other, but neither correlated with ΔOSL (Figure 2). The objective measure, at least within this mental health platform, may therefore be a valuable differentiated source for how mental wellbeing is impacted by therapy in real time.

Objective data were superior to subjective data for therapy recommendation algorithm training

Algorithm performance was evaluated using linear regression analysis, within-subject ANOVA, and ten-fold bootstrapping to compare the efficacy of algorithmically selected (AS) sessions against user selected (US) sessions, using the same wellbeing measure (ΔOSL, ΔSRS, or ΔSRM) for both training and testing.

When ΔOSL or ΔSRM data were used, AS sessions demonstrated increasing efficacy relative to US sessions as a function of training set size, as revealed by linear regression analysis (ΔOSL: $F(1, 16)=5.37$, $P=.034$; ΔSRM: $F(1, 16)=3.69$, $P=.048$; Figure 3A,C). No effect of training set size was observed when ΔSRS date were used ($F(1, 16)=0.046$, $P=.83$, Figure 3B).

Within-subject ANOVA comparisons allowing the algorithm to “learn” from 15 training sessions revealed that ΔOSL training data informed an algorithm that recommended significantly more efficacious sessions when compared with US sessions ($F(1, 66)=8.22$, $P=.005$; Cohen’s d=0.17; Figure 3D). In contrast, the same comparisons when self-reported data were used did not reach significance (ΔSRS: $F(1, 66)=1.24$, $P=.092$; Cohen’s d=0.095; ΔSRM: $F(1, 66)=2.21$, $P=.120$; Cohen’s d=0.041; Figure 3E,F).

Ten-folded bootstrapping results, consistent with the ANOVA, suggested an absence of over-sampling (Figure 3G-I).

Patients gravitate toward more efficacious sessions

The purpose of the recommendation system explored in this study is to ensure that therapy sessions delivered to patients are as efficacious as possible. This being the case, we were curious to examine the overlap of AS and US sessions. Post-hoc analysis via an ANOVA using the expected *versus* actual quantity of sessions with overlap (1, 2, 3+) revealed a sharp rise in the number of patients with overlap between AS and US (Figure 4). This rise in overlap quickly exceeded chance (ΔOSL: $F(1, 14)=30.94$, $P<.001$; ΔSRS: $F(1, 14)=19.33$, $P<.001$; ΔSRM: $F(1, 14)=67.17$, $P<.001$) and followed a quadratic relationship as revealed by regression analysis (ΔOSL: $F(2, 22)=204.1$, $P<.001$; ΔSRS: $F(2, 22)=23.98$, $P<.001$; ΔSRM: $F(2, 22)=51.55$, $P<.001$). These results may hint that users intuitively know – perhaps subconsciously – which sessions are providing benefit.

Discussion

In the current study, we both trained and evaluated a recommendation algorithm with an objective measure of psychological stress. We also performed the identical analyses using two EMAs. We compared the efficacy of algorithmically recommended sessions against sessions selected by patients themselves. Our results indicate that a precision psychiatry recommendation algorithm, when trained with an objective measure of psychological stress, can faithfully predict what forms of therapy will be more efficacious on average than the therapy patients will choose for themselves. The same clinical benefit of recommendations was not observed when the algorithm was trained with subjective EMA data.

This is the first prospective study to our knowledge that leverages a rapid and scalable objective measure of wellbeing to train and evaluate a precision psychiatry therapy recommendation algorithm. The results demonstrate the predictive ability of the approach, and provide a framework for statistical analysis, measurement methodology, and algorithm design in the pursuit of precision psychiatry at a scale that meets global demand. Several interesting, and some surprising, observations emerged during our analysis.

First, patients garnered a mean benefit from engaging with content on the platform. This observation is an important real-world corroboration of previous research [21-24,26,27,30-33]. Without a strong foundation of efficacy from mobile delivery of asynchronous and on-demand therapy, recommendation algorithms embedded within these products are of no value to patients or health systems.

Second, the deltas in subjective *versus* objective data appear to measure separate therapy-response phenomena (Figure 2). Additional support for divergence between objective and subjective data deltas was also revealed in finding that users consistently self-reported larger changes to their wellbeing than what was revealed by objective data. Similar divergence between objective and subjective measures of mental wellbeing has been reported by others [39] and may arise in part from the susceptibility of self-reported data to biases of recall, confirmation, response, and expectation. While EMAs may circumvent recall bias, the biases of confirmation, expectation, and response likely persist. In addition, EMAs rely on a precise awareness of one’s own immediate psychological states, which likely varies between (and within) individuals. Meanwhile, convergence of the EMAs suggests the potential for diminishing returns when collecting multiple sources of self-reported data, at least when compared to a divergent measure like the psychological stress DNN. Thus, irrespective of the mechanism underlying divergence between objective and subjective wellbeing data deltas, their dissociation may serve as an independent rationale for including objective data on mental wellbeing in the practice of psychiatry.

Third, objective data on psychological stress captured immediately before and after therapy sessions were sufficient to train an algorithm to prospectively predict what therapy sessions would be more efficacious for a patient, compared with the content patients would choose for themselves (Figure 3A,D,G). This finding is a realization of the general potential for objective measures of wellbeing in psychiatry. Beyond facial biomarkers, other methods, such as biomarker analysis of speech [40], will likely be critical. Being less susceptible to biases and independent from interoception accuracy, objective data could be more sensitive to factors important for predicting wellbeing outcomes [41,42]. The finding that an objective measure of stress can train an algorithm to select ideal content for patients is especially exciting considering its rapidity of collection (30 s) and potential for broad adoption within healthcare since the software runs on hardware most patients have with them all day (smartphone).

Fourth, subjective data in this study were not sufficient to train the recommendation algorithm to an extent that its predictions were statistically superior to self-selected sessions (Figure 3E,F). While other studies have reported success with recommendation algorithms trained with subjective data e.g., [34], these studies generally compared algorithm recommendations against randomly selected content (instead of the more meaningful patient-selected content, as was the case in our analysis). As discussed above, the limitations of subjective data for algorithm training are likely routed in the multifarious, well-described biases. In particular, an expectation bias was easily observable in our dataset (Figure 3D-F; green markers), and may have impaired the utility of self-report data for training the recommendation algorithm. It is also possible that subjective data have greater session-to-session variability (more noise), giving rise to a less consistent training dataset.

Fifth, with respect to self-reports, when the recommendation algorithm was trained with ΔSRM data, algorithm predictiveness improved as a function of the number of sessions in the training dataset (Figure 3C). No such relationship was found for ΔSRS (Figure 3B). This result may be due to the direct *versus* implicit nature of the stress slider *versus* mood board EMAs, respectively. The conclusion, if these results can be generalized, is that the less subjectivity inherent to a given measure, the better it may be for recommendation algorithm training.

Sixth, there was a surprising overlap of AS and US sessions beyond what would be expected by chance. This finding may suggest that patients intuitively gravitate towards more efficacious therapy regimens. While this is surely a comforting finding for both healthcare professionals and patients, we have little insight into why this overlap occurs. We did note that patients completed identical sessions multiple times, potentially signaling a preference to repeat sessions that led to greater feelings of wellbeing. Even if patients were not consciously aware of the subtle response differences to various forms of content, they nevertheless appear to act as their own recommendation algorithm. More simply put, the sessions patients “like” tend to also be more beneficial. The power of an accurate automated recommendation algorithm lies in improving upon this process regardless of an individual’s ability to self-select efficacious content or gain access to a clinical expert, and thereby rendering the delivery of appropriate care more equitable across patient populations.

The work here demonstrates an opportunity for objective measures of wellbeing, such as psychological stress when rapidly and accurately obtained via a smartphone camera [14], to meaningfully inform precision psychiatry therapy recommendation algorithms. There are additional, broader benefits for psychiatric practice. In particular, removing bias will facilitate higher accuracy and standardization, ensuring a more uniform approach to diagnosis and treatment. Similarly, with the ability to observe measurable impact in real time, patients may become more engaged in beneficial treatment regimens. Objective measures may also lead to earlier detection, allowing intervention before issues become so severe, they are undeniable or result in serious harm.

More broadly, using objective wellbeing data may reduce the stigma associated with an “invisible” self-reported condition and place mental health on the same level as physical health. When readily accessible for clinicians, objective data obtained remotely before a consultation may also reduce the time clinicians spend collecting patient data in person, freeing up time for clinicians to better understand and connect with their patients on a personal level. Objective data will also facilitate the proper inclusion of mental health into the broader healthcare economy, given that the payors of healthcare services are reluctant to cover costs that lack standardization and have susceptibility to “gaming”. Finally, as we show here, objective data will facilitate better personalization of mental healthcare as recommendation methods continue to improve.

Our own future work within precision psychiatry may explore the impact of additional weighted scalers for the current recommendation algorithm, as well as other algorithm designs, such as indication-specific content filtering. For example, patients with major depressive disorder may respond best to consistency with the therapy guide, while patients with generalize anxiety disorder may be particularly influenced by the degree of silence afforded to them during a given therapy session. Thus, there is room to introduce additional content classifiers for consideration by the algorithm. Separately, efficacy of psychotherapy may be moderated by various demographic factors, including cultural background [43]. Mindfulness training and meditation courses are indeed often adapted for specific clinical populations [44-46], age groups [47,48], and cultural groups [49-51]. Incorporating diagnoses, medical history, and/or demographic factors into the algorithm design may lead to further improvement in its predictions. However, these types of data can also introduce undesired biases into the algorithm that counteract our intent to balance functionality across individuals. Caution will therefore be required, keeping in mind that content filters may not impart any net benefit.

Future analysis may also examine additional objective measures of wellbeing beyond psychological stress. In particular, biomarkers from microexpressions and facial blood distribution could be leveraged to produce accurate and objective measures of affect across the domains of valence and arousal. Once validated, multiple objective measures may prove synergistic in their ability to capture the psychological profile of patients as they engage with therapy, and tailor treatment recommendations with ever-increasing accuracy.

There are a few limitations to this study. First, while our DNN quantifies stress from HRV with unprecedented precision [14], the use of facial biomarkers as the ultimate data source requires patients to actively engage with each measurement by way of a mobile selfie scan, ultimately limiting the length of the analysis period, and complicating data capture during therapy (as opposed to before and after). Additional technology developments are underway to facilitate continuous passive facial biomarker data capture, without direct and continuous engagement from the patient. Moreover, the increased convenience and accessibility incurred from avoiding wearables, along with the additional data that can be captured from the human face compared to the finger or wrist, makes the facial biomarker method preferable for precision psychiatry applications at a global level. Second, the dataset used in the current study was restricted to English therapy sessions and therefore most of these patients likely use English as a first language. This constraint was necessary to ensure homogeneity amongst the therapeutic content. While there are no immediately obvious reasons why the results with English content would not generalize to other languages, the restriction does create a bias to a certain patient demographic. It would be prudent and worthwhile to replicate the analyses with non-English therapy sessions once the prerequisite volume of data are at hand.

While our analysis focused on the clinical utility of leveraging objective *versus* subjective data on patient wellbeing for applications within precision psychiatry, the study design and clinical implementation of the technology also demand that the recommendation algorithm training data are obtained rapidly and at scale. Rapidity is essential since the likelihood of patients completing assessments is reduced as a direct function of how long the assessment takes. Meanwhile, scalability is essential to maximize the breadth of patients that can meaningfully benefit from the technology. Wearables and other medical hardware are major barriers for underserved populations, especially in lower socioeconomic regions of developed nations and throughout lower-middle income countries

(LMICs) and lower income countries (LICs). The general theme, of course, is removing barriers and increasing convenience.

Globally, more than 10% of people today need some form of mental healthcare, and more than 50% of us will require mental healthcare at some point in our lives [52]. The current study demonstrates that effective and accessible precision psychiatry can be delivered at scale. We hope our work helps prepare the field to join the rest of medicine in consulting objective data as a regular course of action in the pursuit of optimal patient care. We also hope efficiencies generated via the technology will allow healthcare professional to spend more time with their patients.

In summary, we present evidence that AI-derived objective stress data, when captured through a mobile device front-facing camera before and after asynchronous therapy sessions, can both identify the forms of therapy that are most efficacious for each patient, as well as accurately predict which future therapy sessions will result in the most clinical benefit for each patient. While we feel this finding marks a major step towards precision psychiatry at scale, it is imperative to keep in mind that precision psychiatry recommendation algorithms are, in fact, recommendations. No *in silica* solution, no matter how accurate and powerful, will be able to adequately address all the needs of a patient.

Acknowledgements

Funding for this work was provided by Mobio Interactive Pte Ltd (Singapore) through the salaries supporting the research and development team. Author contributions as follows: HW: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing - review & editing. NASF: Methodology. Data interpretation. Writing - review & editing. BJS: Methodology, Data interpretation, Writing - review & editing, Supervision, Project administration.

Data Availability

The Terms of Use and Privacy Policy afforded to patients using the platform prevents public sharing of the examined dataset, however interested researchers are welcome to contact BJS for collaborations in-line with the promises made by Mobio Interactive Pte Ltd to its patients.

Conflicts of Interest

HW, NASF, and BJS are a Staff Cognitive Neuroscientist, Founding Advisor, and Chief Scientist and CEO of Mobio Interactive Pte Ltd, with each owning 0%, ≈1%, and ≈20% of the company at the time of manuscript submission, respectively. NASF and BJS did not directly contribute to, or have direct influence over, data collection or analysis.

References

1. Torres, E. B., Isenhower, R. W., Nguyen, J., Whyatt, C., Nurnberger, J. I., Jose, J. V., Silverstein, S. M., Papathomas, T. V., Sage, J., & Cole, J. (2016). Toward Precision Psychiatry: Statistical Platform for the Personalized Characterization of Natural Behaviors. *Frontiers in Neurology*, *7*. https://doi.org/10.3389/fneur.2016.00008
2. Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., Baldwin, H., Danese, A., Fazel, S., Steyerberg, E. W., Stahl, D., & Fusar-Poli, P. (2021). Implementing Precision Psychiatry: A Systematic Review of Individualized Prediction Models for Clinical Practice. *Schizophrenia Bulletin*, *47*(2), 284–297. https://doi.org/10.1093/schbul/sbaa120
3. Zanardi, R., Prestifilippo, D., Fabbri, C., Colombo, C., Maron, E., & Serretti, A. (2021). Precision psychiatry in clinical practice. *International Journal of Psychiatry in Clinical Practice*, *25*(1), 19–27. https://doi.org/10.1080/13651501.2020.1809680
4. Colombo, D., Suso-Ribera, C., Fernández-Álvarez, J., Cipresso, P., Garcia-Palacios, A., Riva, G., & Botella, C. (2020). Affect Recall Bias: Being Resilient by Distorting Reality. *Cognitive Therapy and Research*, *44*(5), 906–918. https://doi.org/10.1007/s10608-020-10122-3
5. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., & Mainen, Z. F. (2014). Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, *17*(11), Article 11. https://doi.org/10.1038/nn.3812
6. Dodd-McCue, D., & Tartaglia, A. (2010). Self-report Response Bias: Learning How to Live with its Diagnosis in Chaplaincy Research. *Chaplaincy Today*, *26*(1), 2–8. https://doi.org/10.1080/10999183.2010.10767394
7. Kahneman, D., & Tversky, A. (2012). Prospect Theory: An Analysis of Decision Under Risk. In *Handbook of the Fundamentals of Financial Decision Making: Vol. Volume 4* (pp. 99–127). WORLD SCIENTIFIC. https://doi.org/10.1142/9789814417358_0006
8. Sigmon, S. T., Pells, J. J., Boulard, N. E., Whitcomb-Smith, S., Edenfield, T. M., Hermann, B. A., LaMattina, S. M., Schartel, J. G., & Kubik, E. (2005). Gender Differences in Self-Reports of Depression: The Response Bias Hypothesis Revisited. *Sex Roles*, *53*(5–6), 401–411. https://doi.org/10.1007/s11199-005-6762-3
9. Bardwell, W. A., & Dimsdale, J. E. (2001). The Impact of Ethnicity and Response Bias on the Self-Report of Negative Affect1. *Journal of Applied Biobehavioral Research*, *6*(1), 27–38. https://doi.org/10.1111/j.1751-9861.2001.tb00105.x
10. Lai, L. C. H., Cummins, R. A., & Lau, A. L. D. (2013). Cross-Cultural Difference in Subjective Wellbeing: Cultural Response Bias as an Explanation. *Social Indicators Research*,

114(2), 607–619. https://doi.org/10.1007/s11205-012-0164-z

11. Robinaugh, D. J., Brown, M. L., Losiewicz, O. M., Jones, P. J., Marques, L., & Baker, A. W. (2020). Towards a precision psychiatry approach to anxiety disorders with ecological momentary assessment: The example of panic disorder. *General Psychiatry*, *33*(1), e100161. https://doi.org/10.1136/gpsych-2019-100161
12. Manchia., M., Pisanu, C., Squassina, A., & Carpiniello, B. (2020) Challenges and Future Prospects of Precision Medicine in Psychiatry. *Pharmacogenomics and Personalized Medicine*, 13, 127–140. http://doi.org/10.2147/PGPM.S198225
13. Khan, M., Fernandes, G., Sarawgi, U., Rampey, P., & Maes, P. (2019). PAL: A Wearable Platform for Real-time, Personalized and Context-Aware Health and Cognition Support. *arXiv*, (arXiv:1905.01352). https://doi.org/10.48550/arXiv.1905.01352
14. Al-Jebrni, A. H., Chwyl, B., Wang, X. Y., Wong, A., & Saab, B. J. (2020). AI-enabled remote and objective quantification of stress at scale. *Biomedical Signal Processing and Control*, *59*, 101929. https://doi.org/10.1016/j.bspc.2020.101929
15. Franklin, T. B., Saab, B. J., & Mansuy, I. M. (2012). Neural Mechanisms of Stress Resilience and Vulnerability. *Neuron*, *75*(5), 747–761. https://doi.org/10.1016/j.neuron.2012.08.016
16. Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T.P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal*, 16, 1057-1072. http://dx.doi.org/10.17179/excli2017-480
17. Chwyl, A. G., Chung, R., Amelard, J., Deglint, D. A., Clausi, & A. Wong. (2016). SAPPHIRE: Stochastically acquired photoplethysmogram for heart rate inference in realistic environments. *2016 IEEE International Conference on Image Processing (ICIP)*, 1230–1234. https://doi.org/10.1109/ICIP.2016.7532554
18. Rosenfeld, A., Benrimoh, D., Armstrong, C., Mirchi, N., Langlois-Therrien, T., Rollins, C., Tanguay-Sela, M., Mehltretter, J., Fratila, R., Israel, S., Snook, E., Perlman, K., Klainerman, A., Saab, B., Thoburn, M., Gabbay, C., & Yaniv-Rosenfeld, Amit. (2021). Big Data Analytics and AI in Mental Healthcare. *Theory and Practice.* 137-171. https://doi.org/10.1016/B978-0-12-820203-6.00001-1
19. Johnston, B. W., Barrett-Jolley, R., Krige, A., & Welters, I. D. (2020). Heart rate variability: Measurement and emerging use in critical care medicine. *Journal of the Intensive Care Society*, 21(2), 148–157. https://doi.org/10.1177/1751143719853744
20. Faust, O., Hong, W., Loh, H. W., Xu, S., Tan, R.-S., Chakraborty, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Heart rate variability for medical decision support systems: A

review. *Computers in Biology and Medicine*, 145, 105407. https://doi.org/10.1016/j.compbiomed.2022.105407

21. Walsh, K. M., Saab, B. J., & Farb, N. A. (2019). Effects of a Mindfulness Meditation App on Subjective Well-Being: Active Randomized Controlled Trial and Experience Sampling Study. *JMIR Mental Health*, *6*(1), e10844. https://doi.org/10.2196/10844
22. Bergerot, C., Bergerot, P., Molina, L., Lee, D., Philip, E., & Bultz, B. (2022). Psycho-Oncology and the Relevance of a Biopsychosocial Screening Program. *Oncology, 36*, 552–556.
23. Bergerot, C. D., Malhotra, J., Bergerot, P. G., Philip, E. J., Govindarajan, A., Salgia, S., Hsu, J., Matos Neto, J. N., Molina, L. N. M., Vasconcellos, V. F. de, Muddasani, R., Chawla, N. S., Castro, D. V., Chehrazi-Raffle, A., Saab, B., Carlson, L. E., & Pal, S. K. (2022). Prospective assessment of a smartphone-app based mindfulness program for patients with metastatic renal cell carcinoma (mRCC). *Journal of Clinical Oncology*, *40*(6_suppl), 324–324. https://doi.org/10.1200/JCO.2022.40.6_suppl.324
24. Bergerot, C. D., Philip, E. J., Bergerot, P. G., Siddiq, N., Tinianov, S., & Lustberg, M. (2022). Fear of Cancer Recurrence or Progression: What Is It and What Can We Do About It? *American Society of Clinical Oncology Educational Book*, *42*, 18–27. https://doi.org/10.1200/EDBK_100031
25. Hood, P., Ramachandran, M., & Devitt, R. (2022). Growing pains: Lessons learned from a failed mobile mindfulness clinical trial for patients with complex care needs. *Contemporary Clinical Trials Communications*, *26*, 100858. https://doi.org/10.1016/j.conctc.2021.100858
26. Patel, A., Calahong, B., Dhawan, M., Leynes, L. P., Daros, A., Saab, B., Quilty, L. (2023, June 20). Effects of a Mindfulness App for Outpatients Waitlisted for Psychological Interventions [Poster presentation]. 2023 *Society for Digital Mental Health Annual Meeting*, Virtual meeting
27. Subnis, U. B., Farb, N. A., Piedalue, K.-A. L., Speca, M., Lupichuk, S., Tang, P. A., Faris, P., Thoburn, M., Saab, B. J., & Carlson, L. E. (2020). A Smartphone App–Based Mindfulness Intervention for Cancer Survivors: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols*, *9*(5), e15178. https://doi.org/10.2196/15178
28. Garrett, B.M., Tao, G., Taverner, T., Cordingley, E., & Sun, C. (2020) Patients perception of virtual reality therapy in the management of chronic cancer pain. *Heliyon.* 6(e03916). https://doi.org/10.1016/j.heliyon.2020.e03916
29. Ledoux, A.-A., Zemek, R., Cairncross, M., Silverberg, N., Sicard, V., Barrowman, N.,

Goldfield, G., Gray, C., Harris, A. D., Jaworska, N., Reed, N., Saab, B. J., Smith, A., & Walker, L. (2024). Smartphone App–Delivered Mindfulness-Based Intervention for Mild Traumatic Brain Injury in Adolescents: Protocol for a Feasibility Randomized Controlled Trial. *JMIR Research Protocols*, Apr 11;13:e57226. https://doi.org/10.2196/57226

30. Sicard, V., O'Kane, K., Brown, O., Butterfield, L., Kardish, R., Choi, E., Healey, K., Silverberg, N., Smith, A. M., Goldfield, G., Saab, B. J., Gray, C., Goulet, K., Anderson, P., Mackie, C., Roth, S., Osmond, M., Zemek, R., Cairncross, M., & Ledoux, A.-A. (2024). Acceptability, usability, and credibility of a mindfulness-based digital therapeutic for pediatric concussion: A mixed-method study. *DIGITAL HEALTH*, *10*. https://doi.org/10.1177/20552076241248296
31. O'Kane KMK, Otamendi T, Silverberg ND, Choi E, Sicard V, Zemek R, Healey K, Brown O, Butterfield L, Smith A, Goldfield G, Kardish R, Saab BJ, Ledoux AA, Cairncross M. (2024) Development of Therapeutic Alliance and Social Presence in a Digital Intervention for Pediatric Concussion: Qualitative Exploratory Study. *JMIR Formative Research* 2024;8:e49133 https://doi.org/10.2196/49133
32. Bergerot, C. D., Malhotra, J., Bergerot, P. G., Philip, E. J., Castro, D. V., Govindarajan, A., Molina, L. N., Fuzita, W. H., Azeredo, A. C., Mota, A. C., Franca, M. V., Anjos, G. dos, Filho, R. D., de Souza, D. M., Hsu, J., Chawla, N., Chehrazi-Raffle, A., Saab, B., Carlson, L. E., & Pal, S. M. (2023). Improved health-related quality of life (HR-qol) with use of an online mindfulness tool in patients with metastatic renal cell carcinoma (MRCC) receiving immunotherapy (IO). *Journal of Clinical Oncology*, *41*(6_suppl), 660–660. https://doi.org/10.1200/jco.2023.41.6_suppl.660
33. Bergerot, C. D., Bergerot, P. G., Philip, E. J., Malhotra, J., Castro, D. V., Govindarajan, A., Fuzita, W. H., Franca, M. V., Azeredo, A. C., Anjos, G. dos, , de Araujo, R., Hsu, J., Chawla, N., Chehrazi-Raffle, A., Buso, M. M., Saab, B., Carlson, L. E., & Pal, S. M. (2025) Feasibility and acceptability of a mindfulness app-based intervention among patients with metastatic renal cell carcinoma: a multinational study. *The Oncologist.* 30(1):309. https://doi.org/10.1093/oncolo/oyae309
34. Bidargaddi, N., Musiat, P., Winsall, M., Vogl, G., Blake, V., Quinn, S., Orlowski, S., Antezana, G., & Schrader, G. (2017). Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial. *Journal of Medical Internet Research*, *19*(5), e6775. https://doi.org/10.2196/jmir.6775

35. Mohr, D. C., Schueller, S. M., Tomasino, K. N., Kaiser, S. M., Alam, N., Karr, C., Vergara, J. L., Gray, E. L., Kwasny, M. J., & Lattie, E. G. (2019). Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. *Journal of Medical Internet Research*, *21*(8), e13609. https://doi.org/10.2196/13609
36. Cheung, K., Ling, W., Karr, C. J., Weingardt, K., Schueller, S. M., & Mohr, D. C. (2018). Evaluation of a recommender app for apps for the treatment of depression and anxiety: An analysis of longitudinal user engagement. *Journal of the American Medical Informatics Association*, *25*(8), 955–962. https://doi.org/10.1093/jamia/ocy023
37. De Croon, R. D., Houdt, L. V., Htun, N. N., Štiglic, G., Abeele, V. V., & Verbert, K. (2021). Health Recommender Systems: Systematic Review. *Journal of Medical Internet Research*, *23*(6), e18035. https://doi.org/10.2196/18035
38. Sadasivam, R. S., Borglund, E. M., Adams, R., Marlin, B. M., & Houston, T. K. (2016). Impact of a Collective Intelligence Tailored Messaging System on Smoking Cessation: The Perspect Randomized Experiment. *J Med Internet Res*, *18*(11), e285. https://doi.org/10.2196/jmir.6465
39. Goyal, A., Singh, S., Vir, D., & Pershad, D. (2016). Automation of Stress Recognition Using Subjective or Objective Measures. *Psychological Studies*, *61*(4), 348–364. https://doi.org/10.1007/s12646-016-0379-1
40. Kappen, M., Vanderhasselt, M.-A., & Slavich, G. M. (2023). Speech as a promising biosignal in precision psychiatry. *Neuroscience & Biobehavioral Reviews*, *148*, 105121. https://doi.org/10.1016/j.neubiorev.2023.105121
41. Browning, M., Kingslake, J., Dourish, C.T., Goodwin, G.M., Harmer, C.J., & Dawson, G.R. (2019) Predicting treatment response to antidepressant medication using early changes in emotional processing. *Eur Neuropsychopharmacol*, 29(1), 66-75. https://doi.org10.1016/j.euroneuro.2018.11.1102
42. Solomon, Z., Mikulincer, M., & Hobfoll, S.E. (1987) Objective versus subjective measurement of stress and social support: combat-related reactions. *Journal of Consulting and Clinical Psychology*, 55(4), 577-583. https://doi.org/10.1037/0022-006X.55.4.577
43. Hall, G.C., Ibaraki, A.Y., Huang, E.R., Marti, C.N., & Stice, E. (2016) A Meta-Analysis of Cultural Adaptations of Psychological Interventions. *Behavior Therapy*, 47(6), 993-1014. https://doi.org/10.1016/j.beth.2016.09.005
44. Carlson, L. E., Tamagawa, R., Stephen, J., Doll, R., Faris, P., Dirkse, D., & Speca, M. (2014).

Tailoring Mind-Body Therapies to Individual Needs: Patients’ Program Preference and Psychological Traits as Moderators of the Effects of Mindfulness-Based Cancer Recovery and Supportive-Expressive Therapy in Distressed Breast Cancer Survivors. *JNCI Monographs*, *2014*(50), 308–314. https://doi.org/10.1093/jncimonographs/lgu034

45. Paller, K. A., Creery, J. D., Florczak, S. M., Weintraub, S., Mesulam, M.-M., Reber, P. J., Kiragu, J., Rooks, J., Safron, A., Morhardt, D., O’Hara, M., Gigler, K. L., Molony, J. M., & Maslar, M. (2015). Benefits of Mindfulness Training for Patients With Progressive Cognitive Decline and Their Caregivers. *American Journal of Alzheimer’s Disease & Other Dementias®*, *30*(3), 257–267. https://doi.org/10.1177/1533317514545377
46. Wang, X., Li, P., Pan, C., Dai, L., Wu, Y., & Deng, Y. (2019). The Effect of Mind-Body Therapies on Insomnia: A Systematic Review and Meta-Analysis. *Evidence-Based Complementary and Alternative Medicine*, *2019*, e9359807. https://doi.org/10.1155/2019/9359807
47. Redden, J. M., Korol, S. A., & Vassallo, T. (2021). An evaluation of a brief guided mindfulness workshop with older adults: A qualitative study. *Mental Health & Prevention*, *24*, 200215. https://doi.org/10.1016/j.mhp.2021.200215
48. Smart, C. M., Segalowitz, S. J., Mulligan, B. P., Koudys, J., & Gawryluk, J. R. (2016). Mindfulness Training for Older Adults with Subjective Cognitive Decline: Results from a Pilot Randomized Controlled Trial. *Journal of Alzheimer’s Disease*, *52*(2), 757–774. https://doi.org/10.3233/JAD-150992
49. DeLuca, S. M., Kelman, A. R., & Waelde, L. C. (2018). A Systematic Review of Ethnoracial Representation and Cultural Adaptation of Mindfulness- and Meditation-Based Interventions. *Psychological Studies*, *63*(2), 117–129. https://doi.org/10.1007/s12646-018-0452-z
50. García-Campayo, J., Demarzo, M., Shonin, E., & Van Gordon, W. (2017). How Do Cultural Factors Influence the Teaching and Practice of Mindfulness and Compassion in Latin Countries? *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01161
51. Woods-Giscombé, C. L., & Gaylord, S. A. (2014). The Cultural Relevance of Mindfulness Meditation as a Health Intervention for African Americans: Implications for Reducing Stress-Related Health Disparities. *Journal of Holistic Nursing*, *32*(3), 147–160. https://doi.org/10.1177/0898010113519010
52. World Health Organization. (n.d.). *Mental health.* World Health Organization. https://www.who.int/health-topics/mental-health#tab=tab_1

Abbreviations

AI: Artificial intelligence

AS: Algorithmically selected

DNN: Deep neural network

DTx: Digital therapeutic(s)

EMA: Ecological momentary assessment

GDPR: General Data Protection Regulation

HIPAA: Health Insurance Portability and Accountability Act

HRV: Heart rate variability

ΔOSL: Objective stress level (delta)

PHIPA: Personal Health Information Protection Act

US: User selected

rPPG: Remote photoplethysmography

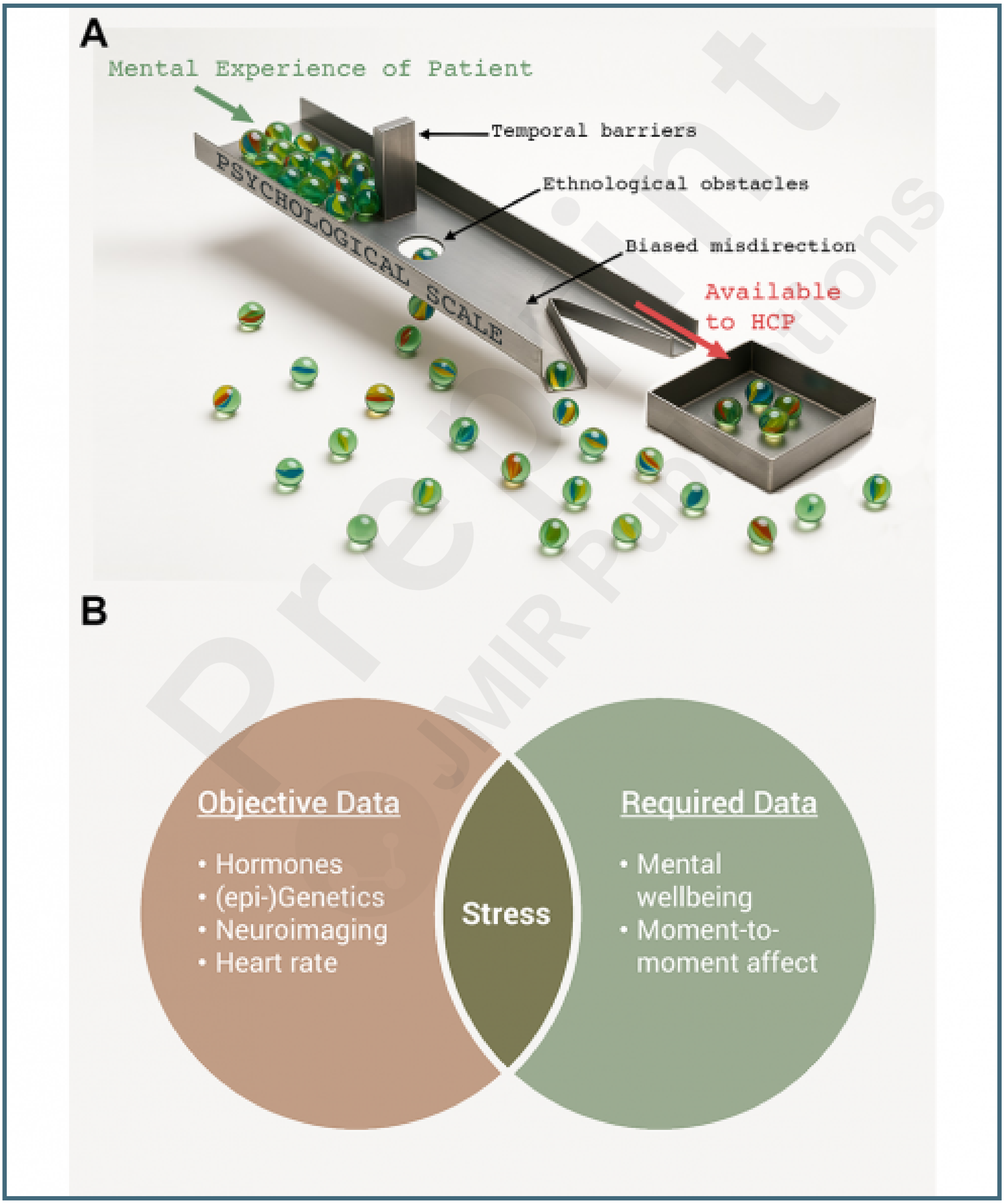
ΔSRM: Self-reported mood (delta)

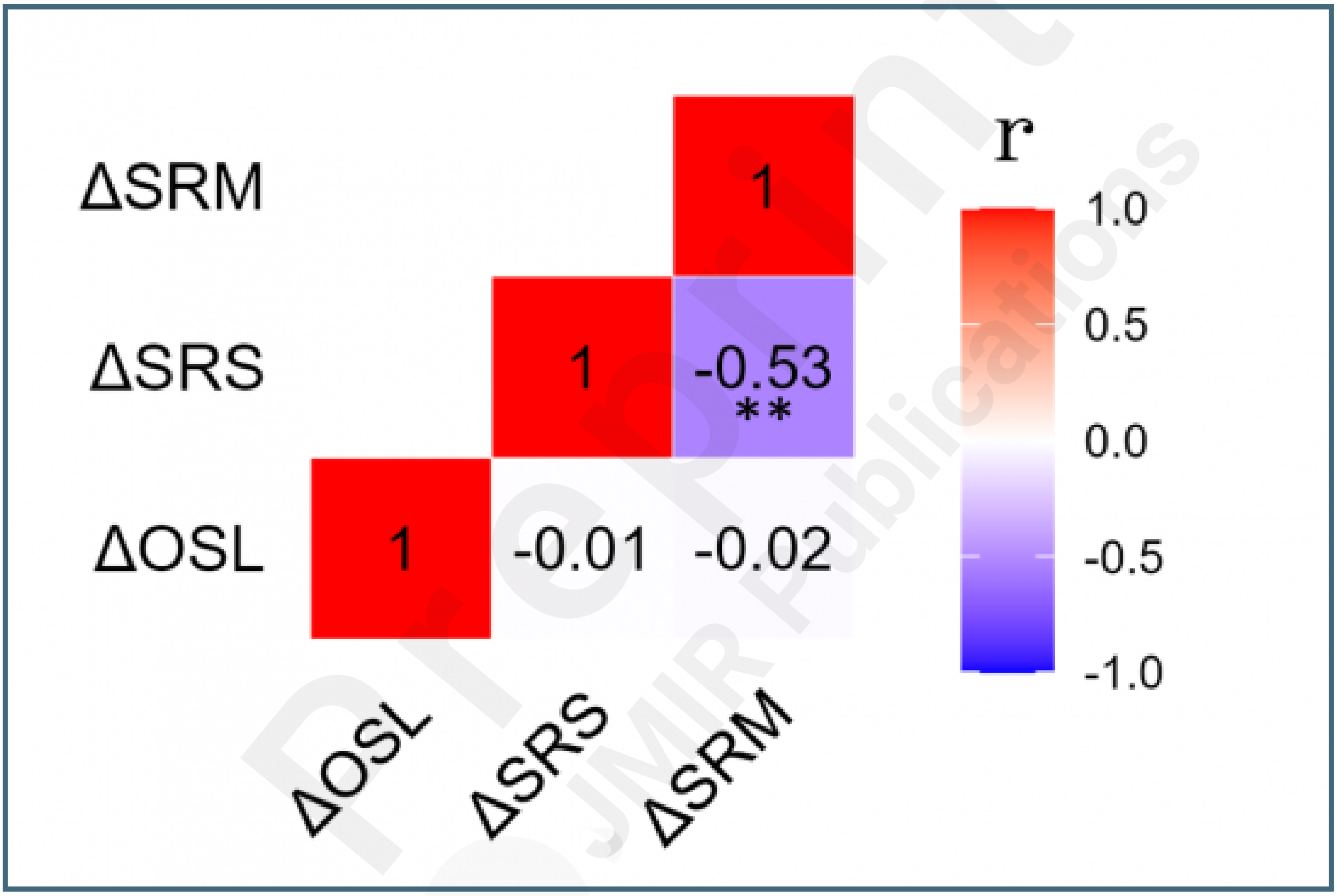
ΔSRS: Self-reported stress (delta)

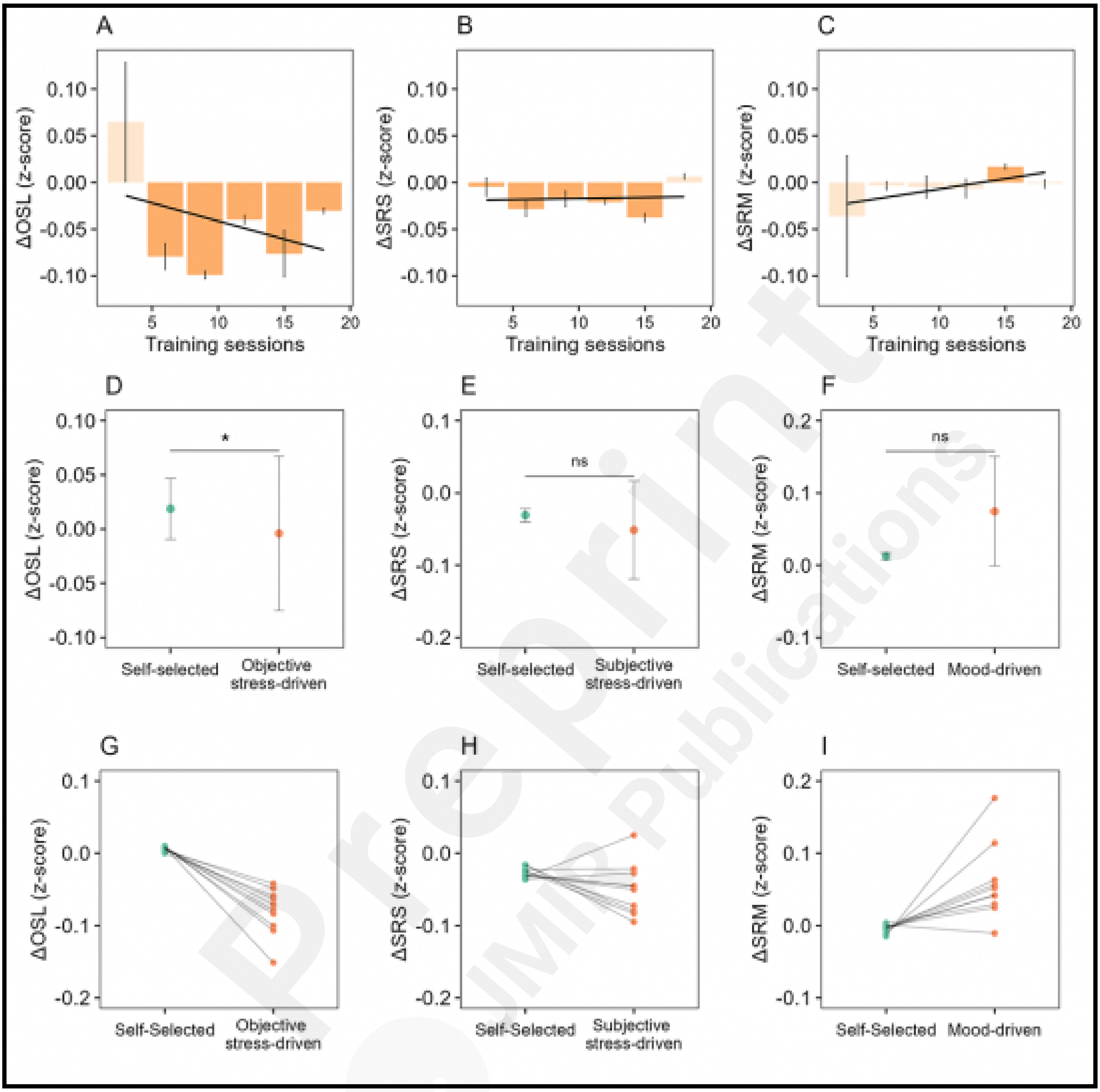
Supplementary Files

Figures

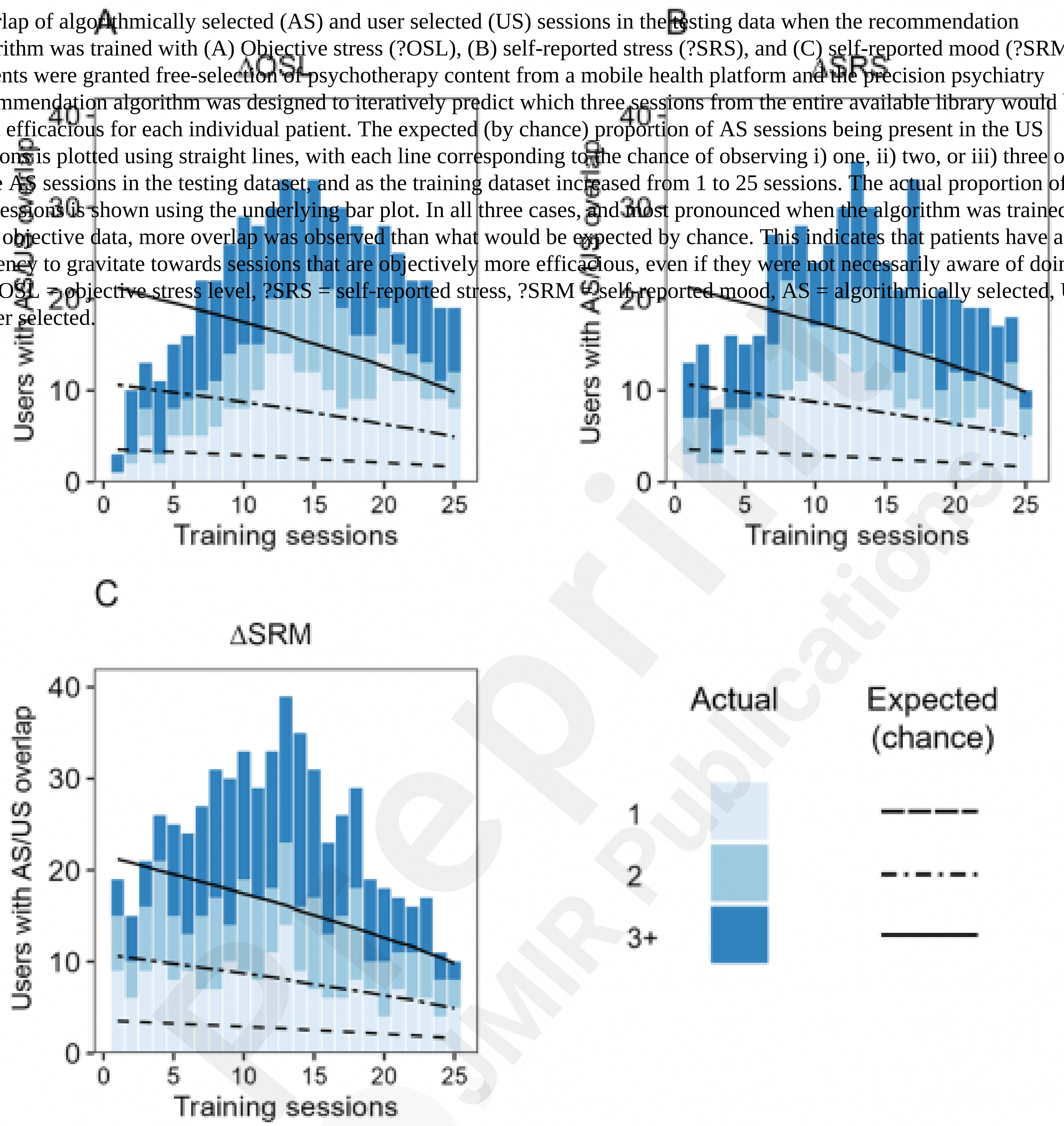
Illustrating known problems of common data sources in psychiatry. (A) Significant limitations inherent to psychological scales: i) temporal barriers, ii) ethnological obstacles, and iii) biased misdirection. While validated scales will remain integral to the diagnosis of mental health conditions in the near-term, subjective scales have less to offer precision psychiatry. (B) Venn diagram of: i) quantifiable/objective data and ii) desired/required data in precision psychiatry. Objective measures often utilized in precision psychiatry (left), such as neuroimaging and multi-omics, have little overlap with the measures that are needed to assess treatment efficacy (right). We propose that an unbiased AI prediction of psychological stress based on heart rate variability may represent an early data source that is both objective and required.




ΔSRM
ΔSRS
ΔOSL
1
1
-0.53
**
1
-0.01
-0.02
ΔOSL
ΔSRS
ΔSRM
r
1.0
0.5
0.0
-0.5
-1.0


A
B
C
ΔOSL (z-score)
ΔSRS (z-score)
ΔSRM (z-score)
Training sessions
D
E
F
*
ns
ns
Self-selected
Objective stress-driven
Subjective stress-driven
Mood-driven
G
H
I
Self-Selected

Overlap of algorithmically selected (AS) and user selected (US) sessions in the testing data when the recommendation algorithm was trained with (A) Objective stress (?OSL), (B) self-reported stress (?SRS), and (C) self-reported mood (?SRM). Patients were granted free-selection of psychotherapy content from a mobile health platform and the precision psychiatry recommendation algorithm was designed to iteratively predict which three sessions from the entire available library would be most efficacious for each individual patient. The expected (by chance) proportion of AS sessions being present in the US sessions is plotted using straight lines, with each line corresponding to the chance of observing i) one, ii) two, or iii) three or more AS sessions in the testing dataset, and as the training dataset increased from 1 to 25 sessions. The actual proportion of AS sessions is shown using the underlying bar plot. In all three cases, and most pronounced when the algorithm was trained with objective data, more overlap was observed than what would be expected by chance. This indicates that patients have a tendency to gravitate towards sessions that are objectively more efficacious, even if they were not necessarily aware of doing so. ?OSL = objective stress level, ?SRS = self-reported stress, ?SRM = self-reported mood, AS = algorithmically selected, US = user selected.



TOC/Feature image for homepages

Psychiatry, unlike all other fields of medicine, focuses on subjective measures as standard practice. While there is a strong trend towards including objective data from genetic and health records analysis, to make any meaningful progress in psychiatry, the field needs to think of the brain as an organ that can also be measured objectively. Put simply, we need objective data on mental wellbeing, on brain states. We need unbiased data on how patients respond to therapy. Fortunately, these types of data are emerging. The associated manuscript showcases that, in comparison to subjective ecological momentary assessments, objective data on psychological stress captured with a smartphone at scale can rapidly train a precision psychiatry therapy recommendation algorithm to accurately predict what will be the most efficacious therapy for each individual patient.