

Word Count (all inclusive): 5,651

Word Count (main article only): 4,036

**Two-way Messaging Therapy for Depression and Anxiety:
Longitudinal Response Trajectories**

Thomas D. Hull*, M.S.
Teachers College, Columbia University
Talkspace

Matteo Malgaroli*, Ph.D.
New York University School of Medicine

Philippa S. Connolly, Ph.D.
Teachers College, Columbia University
Mount Sinai Beth Israel Medical Center, NY

Seth Feuerstein, M.D., J.D.
Yale University Center for Biomedical and Interventional Technology

Naomi M. Simon, M.D., MSc
New York University School of Medicine

*These authors contributed equally to this work

Corresponding Author:

Thomas D. Hull, M.A., M.S.
Teachers College Columbia University, Talkspace
33 West 60th Street, 8th Floor
New York, NY 10023
(480) 254-8815
tdh732@mail.harvard.edu

Disclosures: Hull is an employee of the messaging platform used for this study. In the past 36 months Dr. Malgaroli has received minor consulting fees from Talkspace, outside the submitted work. Dr. Feuerstein serves as an advisor and/or Board member to Vibrant!, Groop Internet Platform, Athyrium, LP, and Magellan Health.

Abstract

Background: Telemedicine is a strategy for overcoming barriers to access evidence-based psychotherapy. Digital modalities that operate outside session-based treatment formats, such as ongoing two-way messaging, may further address these challenges. However, no study to date has established suitability criteria for this medium.

Methods: A large outpatient sample ($n = 10,718$) engaged in daily messaging with licensed clinicians from a telemedicine provider. Patients consisted of individuals from urban and rural settings in all 50 states of the US, who signed to a telemedicine provider. Using a longitudinal design, symptoms changes were observed during a 12 week treatment course. Symptoms were assessed from baseline every three weeks using the Patient Health Questionnaire (PHQ-9) for depression, and the Generalized Anxiety Disorder (GAD-7) for anxiety. Demographics and engagement metrics, such as word count for both patients and therapists, were also assessed. Growth mixture modeling was used to tease symptoms trajectories, and identify predictors of treatment response.

Results: Two subpopulations had GAD-7 and PHQ-9 remission outcomes (*Recovery* and *Acute Recovery*, 30.7% of patients), while two others showed amelioration of symptoms (*Depression* and *Anxiety Improvement*, 36.9% of patients). Two subpopulations experienced no changes in symptoms (*Chronic* and *Elevated Chronic*, 32.4% of patients). Higher use of written communication, patient characteristics, and engagement metrics reliably distinguished patients with the greatest level of remission (*Recovery* and *Acute Recovery* groups).

Conclusions: Remission of depression and anxiety symptoms was observed during delivery of psychotherapy through messaging. Improvements rate were consistent with face-to-face therapy, suggesting the suitability of two-way messaging psychotherapy delivery. Characteristics of improving patients were identified and could be used for treatment recommendation. These findings suggest the opportunity for further research, to directly compare messaging delivery with a control group of treatment as usual.

Keywords: telemedicine, depression, anxiety, longitudinal design, digital health, messaging therapy, text therapy

Trial Registration: Clinicaltrials.gov Identifier: NCT03699488, Retrospectively Registered October 8, 2018

Background

Anxiety and depression are the leading cause of disability in middle- and high-income nations.¹ Non-pharmacological treatments for these disorders include a variety of evidence-based psychotherapies, which have consistently been found to be effective.²⁻⁴ Nevertheless, access to mental health care is low, leaving many affected individuals untreated.^{3,5,6} Barriers to care can occur at several levels⁷ and include issues such as geographic remoteness, economic or insurance constraints, work or childcare related conflicts, shortage of practitioners, stigmatization, and physical impairment.⁸⁻¹² The need to mitigate inequalities of access to care highlights the opportunity for innovative approaches to enhance treatment delivery.^{13,14}

Telemedicine interventions offer a solution to increase accessibility, capable of overcoming both geographic and mobility barriers, as well as reducing wait times. A large number of studies have shown that therapy delivered via technology platforms can be effective in symptom reduction, across a range of psychiatric diagnoses.¹⁵⁻²⁰ Research on technology-mediated treatment has largely been conducted on its most common medium, live video. Newer forms of synchronous and asynchronous delivery have so far received much less attention as a form for delivering direct clinical care. One promising example is two-way multimedia messaging (MMS, or “texting”), given the wide availability of platforms and familiarity with texting as a form of communication. MMS has been effectively used in

the past as an adjunct to clinical care, occupying the role of a reminder system or symptom tracker focusing on promoting healthy lifestyle behavior and medication adherence.^{9,21} Synchronous MMS, or “live chat,” has also been used to deliver psychotherapy, and was shown to be effective in combination with a primary care provider.¹⁸ The next generation of MMS treatment has piloted asynchronous modes of delivery, in an attempt to reduce scheduling barriers, expedite treatment initiation, and increase access. In this approach, patients are free to message their provider an unlimited amount 24/7, and clinicians respond during pre-identified times each day for at least five days a week. Preliminary evidence suggested that this approach may be an acceptable and potentially effective medium for conducting therapy.^{17,22} However, these findings were limited by relatively small sample sizes and retrospective reporting. Further research with larger samples is needed to provide evidence base for asynchronous MMS. If effective, this modality would have the potential to enable more equitable and accessible care by substantially increasing the scale of telehealth.

The current study extends prior findings on asynchronous messaging interventions in two important ways. First, we evaluate the feasibility of wide-scale implementation of asynchronous messaging for delivering therapy within a large provider network, using a longitudinal naturalistic design reflecting the use of this medium in practice.²³⁻²⁵ Second, the availability of a very large sample enabled the identification of subpopulations within the wider group in order to investigate the

heterogeneity of response to the medium in terms of clinical outcome and patient characteristics. This study was designed with a focus on external validity and relevance to clinical practice in the community.

We investigated patterns of response to this novel treatment delivery, and whether large numbers of patients and therapists would find the communication modality feasible. As a result, we did not examine the effectiveness or efficacy of specific types of psychotherapy. Instead, we measured patterns of utilization and dropout, while establishing predictors of response and suitability for the modality. Of particular interest was the number of words and messages sent by patients and by therapists, as well as patient characteristics and baseline severity. We hypothesized that patients who sent more words on average would report above average outcomes, and that clinicians who sent more words than average would have patients reporting above average outcomes. We report the outcome trajectories that emerged as important exploratory findings for establishing different rates of change and characteristics for this medium.^{23,26}

Methods

Setting

The study was conducted with a telemedicine platform (Talkspace) used by independently practicing, licensed therapists in the United States. The platform is accessible through internet search, through Employee Assistance Programs, and as a behavioral health benefit through some

individual insurances. Patients first meet with an intake clinician through a live messaging system to conduct a brief, standardized intake to identify the presenting complaint, patient treatment history, and the patient's provider preferences. This information informs a matching algorithm that prioritizes and presents three providers with the desired characteristics for the patient to choose among. Once a clinician is chosen, the provider is alerted, and the patient is immediately introduced to the messaging "room" where treatment takes place. Patients complete a self-report baseline assessment and the provider walks them through the informed consent and emergency contact process after which treatment can begin. Observations in this study include data collected as part of organizational quality assurance and program management processes between January 1, 2016 and February 1, 2018. All patients and clinicians give written consent to the use of their data in a de-identified, aggregate format as part of the user agreement before they begin using the platform. Study procedures were approved as exempt by the institutional review board at Teachers College, Columbia University (15-426).

Participants

Patients

Participants were individuals who presented with a chief complaint of anxiety or depression, were seeking treatment through the service, and who completed at least one PHQ-9 and/or GAD-7. Inclusion criteria consisted of: (1) being English speakers in the United States, (2) between the ages of 18

and 65, (3) having regular internet or cellphone access, (4) receiving a depression or anxiety diagnosis from their assigned licensed mental health provider based on a clinical intake and live messaging or video-based interview, as recorded in the electronic medical record with ICD-10 codes, (5) scoring 10 or higher on the PHQ-9 and/or GAD-7. Exclusion criteria consisted of current or past diagnoses of: (1) bipolar disorder, (2) any schizophrenia spectrum and psychotic disorder, or psychotic features, (3) any medical or neurological condition that would better account for the symptoms, (4) substance or alcohol use disorder (5) any condition requiring hospitalization; or (6) suicidal thoughts and/or behavior sufficient to be marked a “Yes” on any of questions three through six (at least thoughts about a potential suicide method), on the Columbia Suicide Severity Rating Scale Lifetime-Recent Screen,²⁷ requiring a more intensive level of care that interrupted treatment on the platform. 23,901 patient records were reviewed with these criteria; the final sample consisted of 10,718 patients.

Clinicians

Clinicians in the provider network were currently licensed in at least one state, were required to have a Masters degree or above, and had at least three years of post-licensure experience delivering mental health care. Clinicians were matched only to patients where licensure included the patient’s residence. There were a total of 1,599 clinicians - 43.7% of whom reported five to nine years of post-licensure experience, and 36.5%

reporting ten or more years of experience. Eighty-eight percent (88.0%) were female. Providers had a mean age of 40 (SD = 10.04) years, and they reported a mix of orientations as part of their provider profile: 62.5% cognitive-behavioral treatment, 22.5% an integrative or eclectic approach, 8% primarily psychodynamic, and 7% third-wave cognitive behavioral interventions.

Methods and Procedures

Intervention

Clinicians and patients asynchronously exchanged text-, audio-, and video-based messages using a secure, HIPAA-compliant platform accessible on mobile devices and on desktop computers. Patients could freely send messages at any time without limit, and all messages were stored for the clinician when they returned to review the message history. Therapists responded to messages from their patients at least once a day, five days a week. Clinicians were expected to adhere to all reporting, professional, and ethical standards for their respective fields, and appropriate referrals were provided for patients judged to need a higher level of care.

The number of words exchanged between therapists and patients is automatically counted as meta-data by the platform, and was used as a proxy to quantify the extent of therapeutic interaction through the asynchronous messaging medium. Raw counts of words sent by clinicians and patients were used in supplementary analyses. Raw counts of the

number of audio and video messages sent by each party were also analyzed.

Assessments

Patients were assessed for depression and anxiety symptoms at baseline and then every three weeks for the duration of treatment, or until the patient opted to stop receiving assessments. Assessments are introduced to patients as an important aspect of their care that facilitates goal setting and to track progress. In this study, five assessments from baseline to week 12 were analyzed, including: Baseline, Week 3, Week 6, Week 9, and Week 12.

The 9-item Patient Health Questionnaire²⁸ was used to identify the clinical severity of depression. Responses on all items were given on a 4-point Likert scale (0 = *Not at all* to 3 = *Nearly every day*) with a total maximum score of 24. Scores greater or equal than 10 have been shown to have high sensitivity and specificity as a threshold for clinical depression, or at least moderate depression.^{29,30}

Anxiety symptoms were assessed with the 7-item Generalized Anxiety Disorder questionnaire.³¹ Responses on all items were given on a 4-point Likert scale (0 = *Not at all* to 3 = *Nearly every day*) with a total maximum score of 21. Scores of 10 or above have been shown having high sensitivity and specificity as a clinically significant threshold for at least moderate anxiety.³²

Patients opting to leave the platform were asked to indicate the

reason for leaving. Reasons included feeling better or meeting their goals, having money concerns, not liking the therapy medium, having frustrating technical issues, not liking their therapist, deciding to continue treatment face-to-face, or longer having the time necessary to engage in treatment.

Data Analytic Strategy

Outcome trajectories of anxiety and depression symptoms over the 12 weeks of treatment were analyzed using Latent Growth Modeling (LGM) in Mplus 8.³⁴ LGM is an unsupervised machine learning method to identify groups with heterogeneous outcomes (i.e., such as responders and non-responders) and examine their differences. Compared to traditional average-effects approaches, LGM analyzes patterns of change in the data over time, to determine whether there are subpopulations within the overall group of patients. For example, patients with severe symptoms at baseline who end with low symptoms versus patients that begin and end treatment with a milder symptom presentation. In the current study, LGM also teased patients with changes in both anxiety and depression symptoms, versus those improving in only one of the two conditions. Another advantage to LGM is that once patients have been grouped into different trajectories (or classes), characteristics that are common to each class can be identified (i.e., covariates). For example, patients who share a remission trajectory may be far more likely to be female or engage with treatment more consistently than those in another class. As such, LGM provides much more

information in understanding how large groups of people respond to a specific treatment delivery than simply looking at pre- and post-assessment scores for the entire sample. Covariates of interest in this study included age, education, gender, weeks in treatment, words per week for the therapist and words for week for patients. A more technical description of each step of the statistical procedure is provided in the next section.

Technical Specifications of the LGM

Prior to the analyses, missing values for variables with ~40% or less missingness³⁵ were iteratively imputed by random forests (500 trees, 10 iterations), using the R package `missForest`.³⁶ Examined predictors were imputed while masking clinical and outcome variables, to prevent information leakage. All LGM models were estimated under missing at random assumptions using maximum likelihood estimation.

To concurrently capture changes in both anxiety and depression outcomes, the LGM modeled concurrent changes of PHQ-9 and GAD-7 scores as parallel processing³⁴. Specifically, two sets of distinct intercept, slope, and quadratic growth parameters were assigned to each symptoms measure, estimating separate trajectories of anxiety and depression over five assessments (weeks: 0, 3, 6, 9, and 12). The patients classes were then determined based on joint patterns of PHQ-9 and GAD-7 scores growth. The optimal number of classes was determined comparing nested unconditional LGM with increasing numbers of classes. Variance of the growth

parameters was fixed to zero, to increase delineation of classes. Examined model fit indices included Bayesian Information Criterion (BIC), sample-size adjusted Bayesian Information Criterion (SSBIC), Akaike Information Criterion (AIC), relative Entropy, Lo-Mendell-Rubin-adjusted likelihood ratio test (L-M-R LRT), and bootstrapped likelihood ratio test (BLRT). The best fitting solution was estimated based on model fit indices, as well as explanatory properties of the solution.^{37,38}

After determining the solution with the best relative fit, demographic variables and weeks before treatment dropout (or completion) were nested as covariates in a conditional LGM, to analyze class membership predictions. Categorical data was subsequently converted into binary variables from modal values. Auxiliary 3-step method multinomial logistic analyses for latent class predictors³⁹ were then performed on the conditional model. This approach to latent class logistic regression analyses takes into account measurement error in the most likely class attributions, to estimate the predictive role of quantitative treatment delivery characteristics (i.e., the average number of words per week used by therapists and clients over the course of treatment) in determining group membership. Word counts were log-transformed to improve odds ratio dose-response interpretability.

Results

Sample Characteristics

Patients were between the ages of 18 and 65, with the majority (55.0%) falling between 26-35 years of age. Women were 78.9% of the patient sample, and 74.9% of patients had Bachelor's degrees or higher education level. Table 1 provides the full distribution of demographic and clinical characteristics.

Treatment duration was on average 9.75 weeks (SD=3.16), with 56.24% of the sample completing a 12-week treatment course. Of the 4,690 patients discontinuing treatment before 12 weeks, reason for termination was reported by a subset (N=1471, 34% of drop-out): Better/goal met (53.3%), money concerns (22.2%), did not like the treatment medium (10.1%), went to face-to-face treatment (6.9%), technical issues (3.0%), did not like their therapist (2.7%), and no longer had time (2.0%).

Figure 1 reports overall symptom scores for anxiety and depression at each observation. On average there were 2.30 (SD=1.32) symptom assessments available per patient. Viewed through the framework of reliable and clinically significant change,⁴⁰ 53.03% of the sample reported PHQ-9 score reductions of 5 or more points and fell below the established threshold for probable depression, and 47.78% of the sample reported GAD-7 score decreases of 5 or more points and fell below the established threshold for probable anxiety by their last observation. Treatment engagement, as measured by word count, was an average of 788.21 words (SD=4790.63) per week of treatment generated by patients and 626.93 (SD=3556.61) by therapists.

[Table 1 about here]

[Figure 1 about here]

Outcome Trajectories of Anxiety and Depression

Table 2 shows the relative fit indices for progressive model solutions ranging from one to seven classes for parallel latent growth models of PHQ-9 and GAD-7 trajectories. Each of these classes represent a possible subpopulation of patients that is present within the full sample. Model fit indices indicated improved model fit for an increasing number of classes, but the L-M-R LRT did not approach significance for the 7-class solution. The 6-class model was the best fitting and was also judged to have the highest level of interpretability and theoretical utility. Therefore, the 6-class model was chosen as the optimal solution.

[Table 2 about here]

The best fitting parallel Latent Growth Model is displayed in Figure 2. The model identified six subpopulations that were differed in their change of depression and anxiety symptoms over twelve weeks of treatment. Probability of distinct class membership for each individual participant was high, with values ranging from .75 to .86. The most common class showed improving GAD-7 and PHQ-9 symptoms scores (*Recovery*, 23.7%). This group was characterized by moderate levels of anxiety and mild depression symptoms at baseline, which steadily lowered over treatment and remained

at subclinical levels. A more rapid recovery pattern was displayed by another class (*Acute Recovery*, 7.0%). This class displayed initial high symptomatology for both anxiety and depression, which sharply decreased below clinical thresholds over the course of treatment. Two additional classes presented with high scores for both GAD-7 and PHQ-9; symptoms in these groups only marginally improved (*Chronic*, 22.6%) or remained elevated (*Elevated Chronic*, 9.8%) through therapy. As such, these two subpopulations did not respond to treatment. Two other remaining classes (*Depression Improvement*, 20.0%; *Anxiety Improvement*, 16.9%) presented with moderate symptoms of depression and anxiety, which improved into milder severity over the course of treatment.

[Figure 2 about here]

Predictors of Symptoms Remission

Conditional LGMs analyzed the role of patients' covariates in predicting remission outcomes (*Recovery* and *Acute Recovery* group). Gender, age, education, and treatment length were nested as covariates in the model, while the number of words used by therapist-patient dyads every week were included as predictors using 3-step auxiliary analysis. There were no substantial changes in the shape and proportions of the trajectories from the unconditional solutions.

[Table 3 about here]

Of note, 3-step latent class logistic regression analyses indicated significant differences in the use of the messaging service between trajectory classes. Patients assigned to both Recovery groups were more likely to engage in weekly written communication with their therapist compared to all other classes, while controlling for age, education, gender, and treatment duration. Moreover, the therapists of individuals in non-Recovery groups were more likely to try to engage their patients using written communication, resulting in higher average therapist words generated per week. The *Acute Recovery* class was differentiated from the other *Recovery* group by higher likelihood of treatment adherence and lower education; nevertheless, the two Recovery classes did not significantly differ between them by amount of words used with their therapist or received by their therapists.

Results from the multinomial logistic regression analyses indicated that, when compared to the recovery groups, all other patients had lower treatment durations and thus were less likely to complete treatment. When specifically compared to the largest *Recovery* group, all other classes had less likelihood of having a college degree or higher - with the exception of the anxiety improvement group. This latter group also was more likely to have higher education levels than the patients in the other *Acute Recovery* group, which in turn had individuals more likely to have at least a bachelor

degree and to be women compared to the *Elevated Chronic* group. Table 3 reports the full estimates and their confidence intervals.

Discussion

This study examined outcome trajectories following 12 weeks of psychotherapy delivered through asynchronous two-way messaging. The study involved a very large sample of treatment seeking individuals with clinician-reported diagnoses of depression or anxiety, endorsing symptoms in the moderate to severe range. Results showed that depression and anxiety symptoms decreased in the majority of the identified subpopulations (67.6% of the sample), with more than a third reporting symptoms below the established thresholds of the measures (*Recovery* and *Acute Recovery*, 30.7%). Clinically significant symptoms improvements were observed in 47.78% of the sample for GAD-7 and in 53.03% for PHQ-9. Nevertheless, the remaining Chronic groups (32.4% of the total sample) endorsed elevated symptoms throughout treatment. The six identified patient outcome trajectories were distinguished by baseline severity, rates of improvement, education level, treatment adherence, and by number of words generated by therapist and patient. In particular, both Recovery groups were associated with higher written engagement during treatment compared to the other groups. These differences could reflect patient characteristics (e.g., greater motivation and treatment readiness), therapist characteristics (e.g., greater interpersonal skills, warmth, and experience with evidence

supported interventions for the presenting complaint) or a combination of the two. Patients assigned to the Chronic and Improvement groups also tended to receive more messages than both *Recovery* and *Acute Recovery* groups, suggesting that therapists may have had to work harder to keep them engaged in treatment or to work through complex challenges; for patients in the *Chronic* and *Elevated Chronic* groups, psychotherapy could have served to maintain their condition. Importantly, the contrast in the number of words exchanged among groups with different outcomes is potentially instructive about the mechanisms of change in this medium; however, further research is needed to investigate the content and frequency of the messaging based interchanges.

In terms of treatment length, the majority of the sample (59.1%) adhered to treatment for the entirety of the twelve weeks of messaging therapy. In particular, patients in the *Acute Recovery* groups had the highest treatment completion likelihood. Available data on reasons for dropout highlighted that termination in these groups was also due to treatment goal completion (53.3%), an important consideration when gauging the acceptability of any form of treatment.²³ Overall, messaging treatment adherence compared favorably to face-to-face interventions, where the modal number of sessions attended in traditional settings is one, with a median of five sessions.⁴¹ The notable increase in treatment length is likely to be accounted for by the accessibility afforded by two-way messaging, which allows patients to asynchronously communicate with their

therapist whenever is most convenient and from any location. It is also interesting to note that such large proportions of the sample are well-educated (75% with a Bachelor's degree or higher) and female (78%). This is likely an artifact of advertising practices and the channels used to promote adoption of the service, than a statement of suitability for any particular population. Research that is less reliant on treatment seeking individuals looking for a digital service will help to resolve the interesting issue of whether messaging therapy is acceptable to a wide variety of demographic groups.

The current study's was able to capture messaging telemedicine in practice, observing a very wide range of patients across the United States, from both rural and urban settings. Despite these strengths, the study's research design also presented limitations. In particular, examining response trajectories without a control group is limited in its ability to determine the relative effectiveness of messaging therapy, or to control for historical factors and spontaneous remission. However, the LGM analyses address criticisms of regression to the mean and capitalizing on mild cases of depression. In the first case, regression to the mean is a statistical effect within homogenous populations, whereas LGM teases apart and identifies each subpopulation and models change over time for that population⁴¹. In the second case, the LGM identified that while there was a population presenting with mild depression, there were other populations as well that reported greater severity at baseline. These are some advantages of using

LGM with this type of data. Nevertheless, future studies with a control arm and /or an active standard of care comparator, and randomized patient assignment would be an important complement for fully testing the efficacy of treatment delivery on the messaging medium.

Other limitations of this study include missing post-baseline assessments for some patients, and no content analyses to supplement word count for further clarifying subpopulation differences and reasons for response or non-response. Future studies might also consider more extensive assessments such as structured diagnostic interviews, and quality of life and function measures; however, setting up the study as a clinical efficacy trial, depending on the demands, could diminish generalizability. Lastly, the subpopulations identified may only generalize to telemedicine settings that use two-way asynchronous messaging for treatment delivery.

Notwithstanding the limitations above, these results complement findings from early research on treatment via messaging,^{17,22} and extend these findings with a much expanded sample size, the use of longitudinal assessment, and by identifying patient characteristics likely to benefit from therapy in this medium as it is practiced in the field.^{24,25} Further investigation into mechanisms of therapeutic action could greatly enrich our understanding of this medium.

Conclusions

There is growing interest in and utilization of modern communication media for treatment interventions, with uptake at a rapid pace ahead of formal research. It is thus critical to examine these forms of intervention to determine their effectiveness and understand who may benefit and be appropriate for this type of care. We identified two response groups for patients with moderate to severe depression and/or anxiety that achieved symptom remission, as well as two groups with symptom improvement. Several prognostic factors and patient characteristics were identified that predicted whether a patient is likely to experience remission that are critical for evaluating the impact of novel treatment modalities. Dropout rates also suggested that this medium affords added convenience that enables patients to continue with treatment for a longer duration than is reported in traditional settings⁴².

Abbreviations

MMS - Multimedia Message Service

PHQ-9 - Patient Health Questionnaire - 9

GAD-7 - Generalized Anxiety Disorder - 7

ICD-10 - International Classification of Diseases -10

HIPAA - Health Insurance Portability Accountability Act

LGM - Latent Growth Modeling

Declarations

Ethics Approval and consent to participate: Study approved by the Teachers College, Columbia University IRB, 15-426. All study participants

provide written consent to the use of their data in de-identified and aggregate format when acknowledging the Terms of Use before treatment begins.

Consent to publish: Not applicable.

Author Contributions

Dr. Malgaroli had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: All authors

Acquisition, analysis, or interpretation of data: MM

Drafting of the manuscript: All authors

Critical revision of the manuscript for important intellectual content: All authors

Statistical analysis: MM

Obtaining funding: TDH

Administrative, technical, or material support: TDH

Study supervision: PSC, SF, MM, NMS

Read and approved manuscript: All authors

Competing Interests: Hull is an employee of the messaging platform used for this study. In the past 36 months Dr. Malgaroli has received minor consulting fees from Talkspace, outside the submitted work. Dr. Feuerstein serves as an advisor and/or Board member to Vibrant!, Groop Internet Platform, Athyrium, LP, and Magellan Health.

Funding/Support: This study was supported by archival data provided by Talkspace.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Role of Funder/Sponsor: The funding source had a minimal role in the design, data collection, and data management as the data provider. The funding source had no role in the analysis, interpretation of the data, or decision to submit the manuscript for publication.

Acknowledgements: Data for this study was provided by Talkspace.

References

1. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*. 2013;382(9904):1575-1586. doi:10.1016/S0140-6736(13)61611-6.
2. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry*. 2000;157:1-45.
3. Young A, Klapp R, Sherbourne C, Wells K. The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry*. 2001;58:55-61. doi:10.1001/archpsyc.58.1.55
4. Elkin I, Shea MT, Watkins JT. National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Arch Gen Psychiatry*. 1989;46(11):971. doi:10.1001/archpsyc.1989.01810110013002.
5. Simon GE, Fleck M, Lucas R, Bushnell DM. Prevalence and predictors of depression treatment in an international primary care study. *Gen Hosp Psychiatry*. 2004;21:1626-1634.
6. Brody DS, Khaliq AA, Thompson TL. Patients' perspectives on the management of emotional distress in primary care settings. *J Gen Intern Med*. 1997;12:403-406. doi:10.1046/j.1525-1497.1997.00070.x.
7. Scheppers E, Van Dongen E, Dekker J, Geertzen J, Dekker J. Potential barriers to the use of health services among ethnic minorities: A review. *Fam Pract*. 2006;23(3):325-348. doi:10.1093/fampra/cmi113.
8. Alleman, JR. Online counseling: The Internet and mental health treatment. *Psychotherapy: Theory/Research/Practice/Training*. 2002;39(2):199-209. doi:10.1037/0033-3204.39.2.199
9. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. *Am J Prev Med*. 2009;36(2):165-173. doi:10.1016/j.amepre.2008.09.040.
10. Hollon S, Munoz RF, Barlow DH, et al. Psychosocial intervention development for the prevention and treatment of depression: Promoting innovation and increasing access. *Biol Psychiatry*. 2002;52:610-630. doi:10.1016/S0006-3223(02)01384-7.

11. Nutting PA, Rost K, Dickinson M, et al. Barriers to initiating depression treatment in primary care practice. *J Gen Intern Med.* 2002;17:103-111. doi: 10.1046/j.1525-1497.2002.10128.x.
12. Young KS. An empirical examination of client attitudes towards online counseling. *Cyberpsychol Behav.* 2005;8:172-177. doi:10.1089/cpb.2005.8.172.
13. Comer JS, Barlow DH. The occasional case against broad dissemination and implementation: Retaining a role for specialty care in the delivery of psychological treatments. *Am Psychol.* 2014;69(1):1.
14. Mohr DC, Hart SL, Howard I, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. *Ann Behav Med.* 2006;32(3):254-258. doi:10.1207/s15324796abm3203_12.
15. Bee PE, Bower P, Lovell K, et al. Psychotherapy mediated by remote communication technologies: A meta-analytic review. *BMC Psychiatry.* 2008;8(1):60. doi:10.1186/1471-244X-8-60.
16. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: A network meta-analysis. *JAMA Psychiatry.* 2019. doi:10.1001/jamapsychiatry.2019.0268.
17. Hull TD, Mahan K. A Study of Asynchronous Mobile-Enabled SMS Text Psychotherapy. *Telemedicine and e-Health.* 2017;23(3):240-247. doi:10.1089/tmj.2016.0114.
18. Kessler D, Lewis G, Kaur S, et al. Therapist-delivered internet psychotherapy for depression in primary care: A randomised controlled trial. *The Lancet.* 2009;374(9690):628-634. doi:10.1016/s0140-6736(09)61257-5.
19. Nelson EL, Barnard M, Cain S. Treating childhood depression over videoconferencing. *Telemed J E Health.* 2003;9:49-55. doi:10.1089/153056203763317648.
20. Reynolds DAJ, Stiles WB, Grohol JM. An investigation of session impact and alliance in internet based psychotherapy: Preliminary results. *Couns Psychother Res.* 2006;6(3):164-168. doi:10.1080/14733140600853617.
21. Militello LK, Kelly SA, Melnyk BM. Systematic review of Text-Messaging Interventions to promote healthy behaviors in pediatric and

adolescent populations: Implications for clinical practice and research. *Worldviews Evid Based Nurs.* 2012;9:66-77. doi:10.1111/j.1741-6787.2011.00239.x.

22. DellaCrosse M, Mahan K, Hull TD. The effect of messaging therapy for depression and anxiety on employee productivity. *J Technol Behave Sci.* 2018:1-5. doi10.1007/s41347-018-0064-4.
23. Baldwin SA, Berkeljon A, Atkins DC, Olsen JA, Nielsen SL. Rates of change in naturalistic psychotherapy: Contrasting does-effect and good-enough level models of change. *J Consult Clin Psychol.* 2009;77(2):203-211.
24. Seligman ME. The effectiveness of psychotherapy. The Consumer Reports study. *Am Psychol.* 1995;50:965-974.
25. Yeh RW, Valsdottir LR, Yeh MW, et al. Parachute use to prevent death and major trauma when jumping from aircraft: Randomized controlled trial. *BMJ.* 2018;363:k5094.
26. Smagula SF, Butters MA, Anderson SJ, et al. Antidepressant response trajectories and associated clinical prognostic factors among older adults. *JAMA Psychiatry.* 2015;72(10):1021-1028.
27. Oquendo MA, Halberstam B, Mann JJ. Risk factors for suicidal behavior: Utility and limitations of research instruments. In: MB. *Standardized Evaluation in Clinical Practice.* 2003:103-130.
28. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr Ann.* 2002;32(9):509-515. doi:10.3928/0048-5713-20020901-06.
29. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606-13. doi:10.1046/j.1525-1497.2001.016009606.x.
30. Kroenke K, West SL, Swindle R, et al. Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: A randomized trial. *JAMA.* 2001;286:2947-55. doi:10.1001/jama.286.23.2947.
31. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med.* 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092.
32. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and

- detection. *Ann Intern Med.* 2007;146:317-325. doi:10.7326/0003-4819146-5-200703060-00004.
33. Obbarius A, van Maasakkers L, Baer L, et al. Standardization of health outcomes assessment for depression and anxiety: Recommendations from the ICHOM depression and anxiety working group. *Qual Life Res.* 2017;26(12):3211-3225.
 34. Muthén LK, Muthén B. *Mplus User's Guide*. 8th ed. Los Angeles, CA: Muthén & Muthén; 1998-2019.
 35. Tang F, Ishwaran H. Random Forest Missing Data Algorithms. *Stat Anal Data Min.* 2017;10(6):363-377. doi:10.1002/sam.11348
 36. Stekhoven DJ, Bühlmann P. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics.* 2011;28(1):112-118.
 37. Jung T, Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Soc Personal Psychol Compass.* 2008;2(1):302-317.
 38. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct Equ Modeling.* 2007;14(4):535-569.
 39. Asparouhov T, Muthén B. Auxiliary variables in mixture modeling: Three-step approaches using M Plus. *Struct Equ Modeling: A Multidisciplinary Journal.* 2014;21(3):329-341.
 40. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59(1):12. doi:10.1037/0022-006X.59.1.12.
 41. Field AP. Homogeneity of Variance. In Salkind, NJ (Ed.), *Encyclopedia of Research Design*. 2010. SAGE Research Methods: Washington, DC.
 42. Gibbons MBC, Rothbard A, Farris KD, et al. Changes in psychotherapy utilization among consumers of services for major depressive disorder in the community mental health system. *Adm Policy Ment Health.* 2011;38(6):495-503. doi:10.1007/s10488-011-0336-1.