ORIGINAL INVESTIGATION

Psilocybin therapy for treatment resistant depression: prediction of clinical outcome by natural language processing

Robert F. Dougherty[1] · Patrick Clarke[1] · Merve Atli[1] · Joanna Kuc[1] · Danielle Schlosser[1] · Boadie W. Dunlop[2] · David J. Hellerstein[3] · Scott T. Aaronson[4] · Sidney Zisook[5] · Allan H. Young[6] · Robin Carhart-Harris[7] · Guy M. Goodwin[1] · Gregory A. Ryslik[1]

Abstract

Rationale Therapeutic administration of psychedelics has shown significant potential in historical accounts and recent clinical trials in the treatment of depression and other mood disorders. A recent randomized double-blind phase-IIb study demonstrated the safety and efficacy of COMP360, COMPASS Pathways' proprietary synthetic formulation of psilocybin, in participants with treatment-resistant depression.

Objective While the phase-IIb results are promising, the treatment works for a portion of the population and early prediction of outcome is a key objective as it would allow early identification of those likely to require alternative treatment.

Methods Transcripts were made from audio recordings of the psychological support session between participant and therapist 1 day post COMP360 administration. A zero-shot machine learning classifier based on the BART large language model was used to compute two-dimensional sentiment (valence and arousal) for the participant and therapist from the transcript. These scores, combined with the Emotional Breakthrough Index (EBI) and treatment arm were used to predict treatment outcome as measured by MADRS scores. (Code and data are available at https://github.com/compasspathways/Sentiment2D.)

Results Two multinomial logistic regression models were fit to predict responder status at week 3 and through week 12. Cross-validation of these models resulted in 85% and 88% accuracy and AUC values of 88% and 85%.

Conclusions A machine learning algorithm using NLP and EBI accurately predicts long-term patient response, allowing rapid prognostication of personalized response to psilocybin treatment and insight into therapeutic model optimization. Further research is required to understand if language data from earlier stages in the therapeutic process hold similar predictive power.

Keywords Psilocybin · Depression · Natural language processing · Emotional breakthrough index · Machine learning · AI

This article belongs to a Special Issue on Psychodelics 2023

✉ Robert F. Dougherty
robert.dougherty@compasspathways.com

1 COMPASS Pathways, London, UK

2 Emory University, Atlanta, GA, USA

3 Columbia University, New York, NY, USA

4 Sheppard Pratt, Baltimore, MD, USA

5 University of California, San Diego, CA, USA

6 King's College, London, UK

7 University of California, San Francisco, CA, USA

Introduction

Psilocybin and depression

Major depressive disorder (MDD) affects one in six adults in their lifetime, is characterized by at least one depressive episode with a duration of 2 weeks or more, and involves clear changes in mood, cognition, and the ability to experience pleasure (Otte et al. 2016). While MDD can often be effectively managed using psychotherapy and/or pharmacological treatments, a significant number of MDD patients do not respond to multiple treatment attempts. These patients form a distinct subgroup referred to as treatment-resistant depression (TRD) (Rush et al. 2006b; Otte et al. 2016).

Patients with TRD suffer severe hardship with an increased risk of impaired cognitive (Vancappel et al. 2021;

Springer

Gregory et al. 2020) and social functioning (Judd et al. 2000; Hirschfeld et al. 2000; Thase and Howland 1994), reduced quality of life (Rathod et al. 2022; Lex et al. 2019), an increased risk of somatic morbidity (Lawrence et al. 2013), especially cardiovascular diseases (van Melle et al. 2004), and suicidality (Souery et al. 2007). In addition, the burden to caregivers, which is often overlooked, includes financial, emotional, physical, and social difficulties related to the care for the depressed person (Tabeleão et al. 2018; van Wijngaarden et al. 2004). Financially, the national cost of TRD in the USA has been estimated to be between \$29 and \$48 billion annually, assuming that 12–20% of all patients with depression have TRD (Mrazek et al. 2014). However, estimates of the prevalence of TRD vary widely in the literature (12–55%) (Wiles et al. 2014; Nemeroff 2007; Rush et al. 2006a; Corey-Lisle et al. 2002) and the actual costs might be higher. Existing options for next step treatment are various but often unsatisfactory leading to a high rate of relapse (Jain et al. 2022; Culpepper et al. 2015; Judd et al. 1998) or serious side effects such as sexual dysfunction, weight gain, and sleep disturbance (Ferguson 2001). The development of alternative therapeutic options for TRD patients with improved efficacy and acceptability remains an important and open challenge within healthcare.

Recently, therapies involving psychedelic treatment have shown positive results in addressing this challenge. Specifically, psilocybin, a tryptamine alkaloid (Passie et al. 2002), has shown potential as an antidepressant in preliminary studies involving patients with MDD (Davis et al. 2021; Carhart-Harris et al. 2021), TRD (Carhart-Harris et al. 2016), and anxiety in life-threatening cancer (Grob et al. 2011; Ross et al. 2016; Griffiths et al. 2016). More recently, the efficacy of COMP360 psilocybin therapy for TRD patients (at a 25 mg dose), was demonstrated in the first adequately powered multi-site, randomized, double-blind phase IIb study (Goodwin et al. 2022). MADRS score change at week 3 was the primary outcome of the study though a large reduction in depressive symptoms was evident on the day following psilocybin dosing and a subsequent durable response was observed in 20% of the 25 mg group at week 12.

As part of the COMP360 treatment, psychological support was provided to ensure the participants' physical and psychological safety and consisted of three phases: preparation, administration, and integration. In the preparation phase, the participant met with a qualified therapist at least three times in order to receive psychoeducational material, prepare for the psychedelic experience, and build trust. During the psilocybin administration session, typically lasting 6–8 h, the same therapist was present and participants were encouraged to attend to their full range of internal experiences. Subsequently, during the integration sessions, the participants were encouraged to reflect on their experiences

and any emerging insights (Tai et al. 2021). These integration sessions were lead by the same therapist who guided the participant through preparation and the psilocybin administration session.

Here we employ a novel measure of the average linguistic sentiment scores of the participant and therapist computed from transcripts of the integration session 1 day after psilocybin administration, together with items from the participant's self-reported Emotional Breakthrough Inventory (EBI) and the participant's psilocybin dose, to predict enduring responses at 3 and 12 weeks post psilocybin administration. Sentiment, specifically valence and arousal in speech, reflects mood states and has been implicated as a classifier of depression (Stasak et al. 2016). Similarly, EBI scores, which are believed to index experiences of emotional release or catharsis, have been shown to predict outcome in a recent psilocybin trial as well (Roseman et al. 2019). Our hypothesis was that the NLP and EBI information collected during the integration session would predict the primary outcome (MADRS scores) measured at 3, 6, 9, and 12 weeks after the psilocybin administration session.

Methods and materials

Study overview

COMP001 was a phase IIb, international multicenter randomized, double-blind study of COMP360, which randomized 233 adults with TRD in 1:1:1 ratio to a single fixed-dose (25 mg, 10 mg, or 1 mg) of COMP360 that was administered with a well-defined psychological support model that emphasizes safety and self-directed enquiry (Tai et al. 2021). The primary efficacy endpoint was change from baseline at week 3 in the Montgomery-Åsberg Depression Rating Scale (MADRS) total score. The primary study results and the details of the protocol have been previously published (Goodwin et al. 2022).

Study participants The analysis was restricted to the clinical trial participants whose sessions were conducted in English and the therapists who provided psychological support to those participants. We note that all therapists were licensed mental healthcare professionals specially trained in the COMPASS psychological support model (Tai et al. 2021), but were not necessarily certified in clinical psychology. As per the study protocol, participants had the option to opt-out of recording their psychological support sessions. From the 233 participants enrolled in the main study, 185 consented to having their psychological support sessions recorded, 107 of these integration sessions were conducted in English. Five of these participants were excluded from the present analysis due to not having a MADRS score recorded at week 3, and

one was excluded due to having a baseline MADRS score that was in the normal range, resulting in 101 participants for the week 3 target outcome prediction. Of these, 11 were missing MADRS scores at 6, 9, and/or 12 weeks, leaving 90 participants for that analysis. More information about both clinical trial participants and the therapists providing the psychological support can be found in Goodwin et al. (2022).

MADRS outcome measure For each trial participant, seven MADRS scores were collected by a remote blinded rater: at baseline, the day after administration, and then at 1, 3, 6, 9, and 12 weeks post-administration. Here we focus on the change from baseline at week 3, and sustained change from baseline through week 12.

A *responder* was defined as a participant whose MADRS score at week 3 post treatment was reduced by at least 50% relative to that participant's baseline MADRS score and a *sustained responder* was defined as a participant with at least a 50% reduction in MADRS score at each of weeks 3, 6, 9, and 12 after treatment.

Predictive model

Our goal was to predict, with a logistic regression model (see Supplementary Material §2 for details on the model fitting), which participants would be week 3 responders and sustained responders through week 12 using information available immediately after treatment. This included NLP metrics from the transcript of the first integration session post psilocybin administration (i.e., 1 day after the administration session), items of the EBI that were measured at the same integration session, and the treatment dose.

The NLP pipeline begins with an audio recording of the integration session following psilocybin administration and ends with four numbers: two to summarize the average sentiment (valence and arousal) of the utterances spoken by participant, and two corresponding numbers for the therapist. Key steps in this process include:

- The creation of the audio recording,
- The transcription of recordings into text,
- Parsing the text into utterances,
- Computing sentiment scores for each utterance,
- And computing session averages of sentiment values.

We describe these steps in detail in the Supplementary Material, see §1.

Integration session sentiment All psychological support sessions were audio recorded with an Apple iPhone (see Supplementary Material §1.1 for full details on audio recording). The audio recordings of the integration sessions were

Table 1 Logistic regression model fit summaries

Model	Pseudo R-squared	Chi-square	p-value	df	N
Week 3	0.514	64.78	1.66e−11	7	101
Sustained	0.435	39.18	1.8e−06	7	90

manually transcribed into dialog text, and each of these transcripts was then parsed into individual "utterances" (see Supplementary Material §1.2 and §1.3 for transcription and utterance parsing details). These utterances were used to estimate session sentiment for the therapist and participant using a novel sentiment model based on a large-language model that was inspired by the two-dimensional scale of emotion described in Russell (1980). Specifically, this model produces valence and arousal scores for each utterance. The valence score corresponds to the positive (pleasantness) or negative (unpleasantness) score typically used in sentiment analysis. The arousal score characterizes the physiological arousal expressed in the utterance and indicates where it lies on the spectrum from bored/calm to tense/alert/excited (see Supplementary Material §1.4 for utterance sentiment, §1.4.1 for the sentiment model and §1.4.2 for the session sentiment score).

Large language models To compute our sentiment valence and arousal scores, we used a zero-shot classifier built on the BART autoencoder and the MNLI language inference dataset (Lewis et al. 2020; Yin et al. 2019; Williams et al. 2018). The classifier is available on the Hugging Face website https://huggingface.co/facebook/bart-large-mnli/tree/main.

Emotional breakthrough inventory The EBI is an eight-item questionnaire introduced by Roseman et al. (2019) which can predict change in wellbeing following psychedelic therapy (Roseman et al. 2019). The EBI summary score is typically computed by taking the average of the eight items. However, we found that summarizing the EBI with a weighted average of the eight items based on the first component of a principal component analysis (PCA) performed slightly better than the simple average (see Supplementary Material §2.1 for details), and thus used the weighted average for our analysis.

Results

Logistic regression model fits

Before exploring cross-validated predictions for the two outcome measures, the logistic regression models created (see Supplementary Material §2.2) were fit to all the data to confirm that the models resulted in an adequate fit. The pseudo R-square values and chi-square significance test for

Fig. 1 Model fit probability distributions estimated using Gaussian kernel density estimation for the week 3 model (top) and the sustained responder model (bottom). We note that the distributions are quite different between the non-responder and responder participants. This difference is used to successfully classify patient response status as shown in §3.2
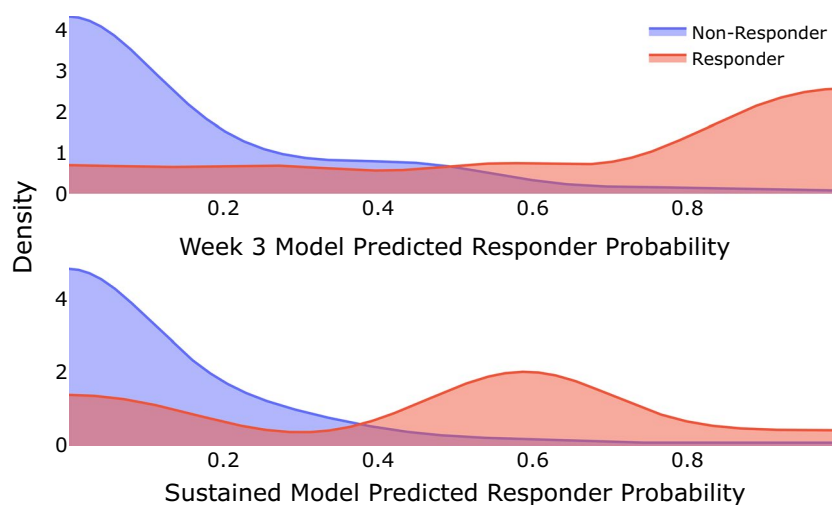
Table 2 Cross-validated model prediction results, where MCC is the Matthews correlation coefficient and f1 is the harmonic mean of precision and recall

Model	Group	Accuracy	MCC	f1	AUC	TN	FP	FN	TP	N
Week 3	ALL	0.851	0.650	0.754	0.877	63	6	9	23	101
Sustained	ALL	0.878	0.610	0.686	0.853	67	5	6	12	90
Week 3	25 mg	0.821	0.641	0.829	0.886	15	3	4	17	39
Sustained	25 mg	0.833	0.639	0.769	0.880	20	3	3	10	36
Week 3	10 mg	0.935	0.746	0.750	0.862	26	0	2	3	31
Sustained	10 mg	0.963	0.693	0.667	0.760	25	0	1	1	27
Week 3	1 mg	0.806	0.380	0.500	0.820	22	3	3	3	31
Sustained	1 mg	0.852	0.250	0.333	0.722	22	2	2	1	27

each of the models are shown in Table 1. These results suggest that the models are well-fit to the data and worthy of further evaluation.

Logistic regression models predict class membership (responder vs. non-responder) probability for each subject. It is useful to inspect the distributions of these class probability predictions to evaluate class separability; the distributions for the models are shown in Fig. 1 and suggest that the two classes should be quite separable.

Cross-validated predictions

The results from the leave-one-out cross-validation for the models described above are shown in Table 2. Leave-one-out cross-validation simulates the case where a model that was trained on existing data is used to predict the responder status of a new participant, and is thus ideal for estimating the real-world performance of a predictive model (see Supplementary Material §2 for details). For each model, results are shown for all participants together, as well as for each of the three treatment groups. A simple bootstrap was also used to compute 95% confidence intervals for the AUC values, which are shown in Table 3. The full receiver operating characteristic plots for these predictions are shown in Fig. 2.

Table 3 AUC values with bootstrapped lower and upper 95% confidence intervals

Model	AUC	Lower	Upper
Week 3	0.877	0.794	0.946
Sustained	0.853	0.747	0.943

Univariate scatterplots

To further understand the relationship between each of the exogenous variables in the predictive models described above and the MADRS scores, we plot each of these variables against the MADRS week 3 change. These scatter plots are shown in Fig. 3, along with a regression line. All six variables are negatively correlated with the week 3 MADRS change, meaning that a higher value on the exogenous variable is associated with a greater MADRS reduction at week 3. This pattern is consistent with the regression model coefficients being generally positive (i.e., higher scores are associated with a higher probability of being a responder). The results are also consistent for the exogenous variables with agreed upon interpretations. For example, as the EBI PCA summary score increases, the MADRS reduction, as expected, increases as well.
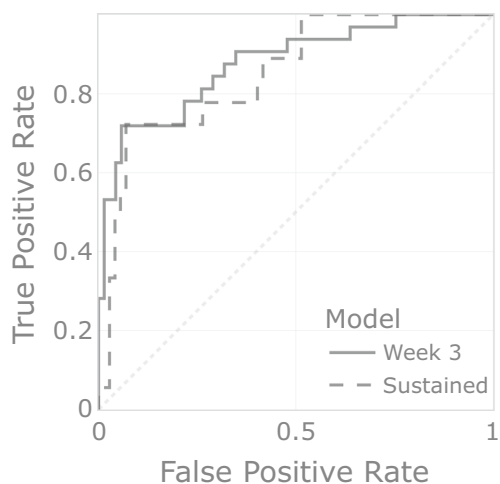
Fig. 2 Receiver operating characteristic curves for the week 3 and sustained responder model predictions

Discussion

For individuals undergoing psilocybin therapy for TRD, response outcomes at 3 and 12 weeks post-dosing could be predicted using the EBI measure or with NLP methods, achieving high accuracy with AUCs ranging from 85 to 88%.

The quality of the interaction between participants and therapists has been shown to be a mediating factor in overall outcome of psilocybin therapy for depression by facilitating a stronger emotional response (Murphy et al. 2022), and the NLP metrics described here capture key aspects of this interaction as it relates to the emotional response. Therapists trained in the COMPASS psychological support model encourage participants to engage with their emotional experiences in the present moment. Engagement with these experiences (without attempting to avoid or control them) is thought to decrease anxiety and enable the participant to have a more therapeutic COMP360 experience. As such it is believed that a stronger emotional experience, as measured by EBI and the NLP metrics, may

Fig. 3 The relationship between the six predictors and week 3 MADRS change ($n = 101$) is shown for therapist arousal (**a**, $r = -0.370$, $p = 0.0001$), therapist valence (**b**, $r = -0.407$, $p = 2e-05$), participant arousal (**c**, $r = -0.397$, $p = 4e-05$), participant valence (**d**, $r = -0.322$, $p = 0.001$), EBI summary score (**e**, $r = -0.556$, $p = 2e-09$) and treatment group (**f**, $r = -0.230$, $p = 0.02$)
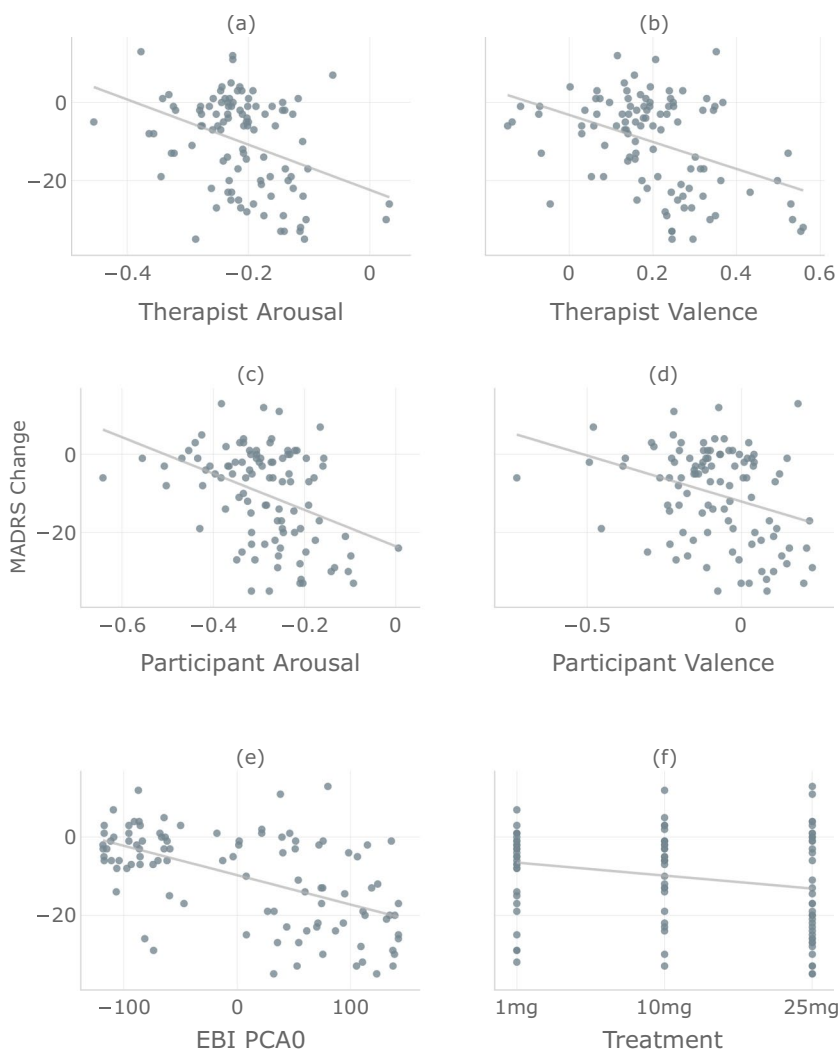
underlie the basis for sustained response to COMP360 treatment. However, we note that while therapists were trained and evaluated in this psychological support model, the lack of fidelity metrics precludes us from evaluating the impact of adherence on outcome and suggests this is a ripe topic for future work.

Moreover, the four NLP metrics and EBI perform about equally well when each is used alone in a partial model (see Supplementary Material §2.3) and statistical performance is improved only moderately when EBI and NLP features are combined into the same model. This suggests that the therapeutic effects may be driven by the same latent variable. Our interpretation is that the fundamental mechanism of COMP360 treatment is to drive emotional breakthrough and a more positive integration session sentiment, which produces a more productive, psychologically insightful session (Peill et al. 2022), and thus results in a positive therapeutic outcome.

We note that the algorithm described in this paper is very fast to train and execute because it relies only upon logistic regression and PCA. As such, a new model may be fit on a standard consumer laptop within seconds of new data becoming available. Moreover, with modern cloud architecture, the transcription and sentiment analysis that forms part of the input to this ML model can be computed in real time. Taken together, this allows model updates to occur at whatever cadence is deemed desirable and allows physician insight into the participant's ultimate response to the treatment immediately after the participant's integration session instead of weeks to months later.

Finally, it is striking that NLP alone gives such a good measure of immediate and early drug effect without recourse to subjective rating scales or even the EBI. It suggests key elements of the participant's mental state can be captured by a passively measured behavioral measure in real time. By utilizing similar information from the preparation and drug administration sessions, we may be able to identify additional signals that improve the accuracy of our predictive model. Furthermore, by using NLP signals from preparation sessions alone, it may be possible to identify which participants are likely to be more responsive to COMP360 therapy prior to the drug administration session. Preparedness may be an important variable for selecting patients for COMP360 therapy, and possibly other psychedelic treatments (Haijen et al. 2018). Additionally, we see future potential for this approach to offer insight into the quality of the psychological support that we assume to be essential for the safe delivery of psychedelic treatments. NLP analysis of the preparation and integration session dialog may also shed light on how treatments may need to be adapted or improved for those who do not respond to an initial treatment.

Limitations of this study include that we were only able to analyze a cohort of 101 participants at 3 weeks and 90 individuals at 12 weeks. While this sample size is more than sufficient for the logistic regression models used in this study, additional subjects from future studies will allow us to further validate the current findings and utilize more advanced machine learning models for prediction such as random forests (Breiman 2001) and neural nets (Mikolov et al. 2013; Vaswani et al. 2017; Devlin et al. 2018; Brown et al. 2020) while reducing the risk of overfitting. Also, because all therapists were trained to follow a specific, well-defined, and non-directive psychological support model (Tai et al. 2021) and not all of the therapists were clinical psychologist, the results may not generalize to other psychedelic therapy contexts. Finally, the burgeoning field of digital biomarkers (Coravos et al. 2019; Jacobson et al. 2019; Cavedoni et al. 2020) may also be integrated to provide signals that correlate with participant response before, during, and after treatment.

In summary, recent advances in utilizing psychedelic treatment for participants with depression show promising results (Goodwin et al. 2022; Davis et al. 2021; Muttoni et al. 2019). In this paper, we demonstrate that we are able to accurately predict treatment response to COMP360 in combination with the associated psychological support model (Goodwin et al. 2022; Tai et al. 2021) using a logistic regression machine learning model that includes NLP metrics from the first integration session, participant's self-response on the EBI scale, and treatment arm. In turn, this ability to predict outcomes has substantial beneficial implications to providers, payers, and participants with respect to future additional treatment or interventions. As additional language and EBI data are collected and combined with digital biomarkers from tools such as smart devices, we expect that this methodology, along with more powerful machine learning models, will further increase our ability to predict participant response to psychedelic treatment. Ultimately, this will allow for more objective, real-time, and personalized care for future patients.

Declarations

References

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901

Carhart-Harris R, Giribaldi B, Watts R, Baker-Jones M, Murphy-Beiner A, Murphy R, Martell J, Blemings A, Erritzoe D, Nutt DJ (2021) Trial of psilocybin versus escitalopram for depression. N Engl J Med 384(15):1402–1411

Carhart-Harris RL, Bolstridge M, Rucker J, Day CM, Erritzoe D, Kaelen M, Bloomfield M, Rickard JA, Forbes B, Feilding A, Taylor D, Pilling S, Curran VH, Nutt DJ (2016) Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. Lancet Psychiatry 3(7):619–627

Cavedoni S, Chirico A, Pedroli E, Cipresso P, Riva G (2020) Digital biomarkers for the early detection of mild cognitive impairment: artificial intelligence meets virtual reality. Front Hum Neurosci 14:245. https://doi.org/10.3389/fnhum.2020.00245

Coravos A, Khozin S, Mandl KD (2019) Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digital Medicine 2(1)14. https://doi.org/10.1038/s41746-019-0090-4

Corey-Lisle PK, Birnbaum HG, Greenberg PE, Marynchenko MB, Claxton AJ (2002) Identification of a claims data "signature" and economic consequences for treatment-resistant depression. J Clin Psychiatry 8:717–726. https://doi.org/10.4088/jcp.v63n0810

COMPASS Pathways (2022) The safety and efficacy of psilocybin in participants with treatment resistant depression. Clinical trial registration NCT03775200, clinicaltrials.gov, submitted: December 11, 2018 (Jan. 2022). https://clinicaltrials.gov/ct2/show/NCT03775200

Culpepper L, Muskin PR, Stahl SM (2015) Major depressive disorder: understanding the significance of residual symptoms and balancing efficacy with tolerability. Am J Med 128(9 Suppl):S1–S15. https://doi.org/10.1016/j.amjmed.2015.07.001

Davis AK, Barrett FS, May DG, Cosimano, Sepeda ND, Johnson MW, Finan PH, Griffiths RR (2021) Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. JAMA Psychiatry 78(5)481. https://doi.org/10.1001/jamapsychiatry.2020.3285

Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/ARXIV.1810.04805

Ferguson JM (2001) SSRI antidepressant medications: adverse effects and tolerability. Primary care companion to the Journal of Clinical Psychiatry 3(1):22–27

Griffiths RR, Johnson MW, Carducci MA, Umbricht A, Richards WA, Richards BD, Cosimano MP, Klinedinst MA (2016) Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: a randomized double-blind trial. J Psychopharmacol 30(12):1181–1197

Gregory E, Torres IJ, Ge R, Blumberger DM, Downar JH, Daskalakis ZJ, Lam RW, Vila-Rodriguez F (2020) Predictors of cognitive impairment in treatment-resistant depression. J Affect Disord 274:593–601. https://doi.org/10.1016/j.jad.2020.05.101

Goodwin GM, Aaronson ST, Alvarez O, Arden PC, Baker A, Bennett JC, Bird C, Blom RE, Brennan C, Brusch D et al (2022) Single-dose psilocybin for a treatment-resistant episode of major depression. N Engl J Med 387(18):1637–1648

Grob CS, Danforth AL, Chopra GS, Hagerty M, McKay CR, Halberstadt AL, Greer GR (2011) Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. Arch Gen Psychiatry 68(1):71–78

Haijen ECHM, Kaelen M, Roseman L, Timmermann C, Kettner H, Russ S, Nutt D, Daws RE, Hampshire ADG, Lorenz R, Carhart-Harris RL (2018) Predicting responses to psychedelics: a prospective study. Front Pharmacol 9:897

Hirschfeld R, Montgomery S, Keller M, Kasper S, Schatzberg A, Möller H, Healy D, Baldwin D, Humble M, Versiani M, Montenegro R, Bourgeois M (2000) Social functioning in depression: a review. Diseases of the Nervous System 61(4)268–275. copyright: Copyright 2017 Elsevier B.V., All rights reserved. https://doi.org/10.4088/JCP.v61n0405

Jain R, Higa S, Keyloun K, Park J, Bonafede M, Tung A, Gillard P, Cutler AJ (2022) Treatment patterns during major depressive episodes among patients with major depressive disorder: a retrospective database analysis. Drugs - Real World Outcomes. https://doi.org/10.1007/s40801-022-00316-4

Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Paulus MP, Kunovac JL, Leon AC, Mueller TI, Rice JA, Keller MB (1998) Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. J Affect Disord 50(2):97–108. https://doi.org/10.1016/S0165-0327(98)00138-4

Judd LL, Akiskal HS, Zeller PJ, Paulus M, Leon AC, Maser JD, Endicott J, Coryell W, Kunovac JL, Mueller TI, Rice JP, Keller MB (2000) Psychosocial disability during the long-term course of unipolar major depressive disorder. Arch Gen Psychiatry 57(4):375–380. https://doi.org/10.1001/archpsyc.57.4.375

Jacobson NC, Weingarden H, Wilhelm S (2019) Digital biomarkers of mood disorders and symptom change. NPJ Digital Medicine 2(1)3. https://doi.org/10.1038/s41746-019-0078-0

Lawrence D, Hancock KJ, Kisely S (2013) The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. BMJ 346. https://doi.org/10.1136/bmj.f2539

Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2020) BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, pp. 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703

Lex H, Ginsburg Y, Sitzmann AF, Grayhack C, Maixner DF, Mickey BJ (2019) Quality of life across domains among individuals with treatment-resistant depression. J Affect Disord 243:401–407. https://doi.org/10.1016/j.jad.2018.09.062

Mrazek DA, Hornberger JC, Altar CA, Degtiar I (2014) A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996–2013. Psychiatric Services 65(8)977–987. pMID: 24789696. https://doi.org/10.1176/appi.ps.201300059

Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. https://doi.org/10.48550/ARXIV.1310.4546

Murphy R, Kettner H, Zeifman R, Giribaldi B, Kartner L, Martell J, Read T, Murphy-Beiner A, Baker-Jones M, Nutt D et al (2022) Therapeutic alliance and rapport modulate responses to psilocybin assisted therapy for depression. Front Pharmacol 12:788155

Muttoni S, Ardissino M, John C (2019) Classical psychedelics for the treatment of depression and anxiety: a systematic review. J Affect Disord 258:11–24. https://doi.org/10.1016/j.jad.2019.07.076

Nemeroff CB (2007) Prevalence and management of treatment-resistant depression. J Clin Psychiatry 68(Suppl 8):17–25

Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, Mohr DC, Schatzberg AF (2016) Major depressive disorder. Nat Rev Dis Primers 2:16065

Passie T, Seifert J, Schneider U, Emrich HM (2002) The pharmacology of psilocybin. Addict Biol 7(4):357–364

Peill JM, Trinci KE, Kettner H, Mertens LJ, Roseman L, Timmermann C, Rosas FE, Lyons T, Carhart-Harris RL (2022) Validation of the Psychological Insight Scale: a new scale to assess psychological insight following a psychedelic experience. J Psychopharmacol 36(1):31–45

Rathod S, Denee T, Eva J, Kerr C, Jacobsen N, Desai M, Baldock L, Young AH (2022) Health-related quality of life burden associated with treatment-resistant depression in UK patients: quantitative results from a mixed-methods non-interventional study. J Affect Disord 300:551–562. https://doi.org/10.1016/j.jad.2021.12.090

Roseman L, Haijen E, Idialu-Ikato K, Kaelen M, Watts R, Carhart-Harris R (2019) Emotional breakthrough and psychedelics: validation of the emotional breakthrough inventory. J Psychopharmacol 026988111985597–026988111985597. https://doi.org/10.1177/0269881119855974

Ross S, Bossis A, Guss J, Agin-Liebes G, Malone T, Cohen B, Mennenga SE, Belser A, Kalliontzi K, Babb J, Su Z, Corby P, Schmidt BL (2016) Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. J Psychopharmacol 30(12):1165–1180

Rush A, Trivedi M, Wisniewski S, Nierenberg A, Stewart J, Warden D, Niederehe G, Thase M, Lavori P, Lebowitz B, McGrath P, Rosenbaum J, Sackeim H, Kupfer D, Luther J, Fava M (2006a) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a star*d report. Am J Psychiatry 163:1905–17. https://doi.org/10.1176/appi.ajp.163.11.1905

Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M (2006b) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 163(11):1905–1917

Russell J (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178

Stasak B, Epps J, Cummins N, Goecke R (2016) An investigation of emotional speech in depression classification. In: Proc. Interspeech 2016 pp. 485–489. https://doi.org/10.21437/Interspeech.2016-867

Souery D, Oswald P, Massat I, Bailer U, Bollen J, Demyttenaere K, Kasper S, Lecrubier Y, Montgomery S, Serretti A, Zohar J, Mendlewicz J (2007) Clinical factors associated with treatment resistance in major depressive disorder. J Clin Psychiatry 68:1062–70. https://doi.org/10.4088/JCP.v68n0713

Tabeleão V, Tomasi E, de Avila Quevedo L (2018) A randomized, controlled trial of the effectiveness of a psychoeducational intervention on family caregivers of patients with mental disorders. Community Ment Health J 54:211–217. https://doi.org/10.1007/s10597-017-0126-7

Tai SJ, Nielson EM, Lennard-Jones M, Johanna Ajantaival RL, Winzer R, Richards WA, Reinholdt F, Richards BD, Gasser P, Malievskaia E (2021) Development and evaluation of a therapist training program for psilocybin therapy for treatment-resistant depression in clinical research. Front Psych 12. https://doi.org/10.3389/fpsyt.2021.586682

Thase ME, Howland RH (1994) Refractory depression: relevance of psychosocial factors and therapies. Psychiatr Ann 24(5):232–240. https://doi.org/10.3928/0048-5713-19940501-09

Vancappel A, Dansou Y, Godin O, Haffen E, Yrondi A, Stéphan F, Richieri R, Molière F, Horn M, Allauze E, Genty J, Bouvard A, Dorey J, Meyrel M, Camus V, Fond G, Péran B, Walter M, Anguill L, Scotto d'Apollonia C, Nguon A, Fredembach B, Holtzmann J, Vilà E, Petrucci J, Rey Etain B, Carminati M, Courtet P, Vaiva G, Llorca P, Leboyer M, Aouizerate B, Bennabi D, El-Hage W (2021) Cognitive impairments in treatment-resistant depression: results from the French cohort of outpatients (FACE-DR). Journal of Affective Disorders Reports 6:100272. https://doi.org/10.1016/j.jadr.2021.100272

van Melle J, de Jonge P, Spijkerman T, Tijssen J, Ormel J, van Veldhuisen D, van den Brink R, van den Berg M (2004) Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis. Psychosom Med 66(6):814–822

van Wijngaarden B, Schene AH, Koeter MW (2004) Family caregiving in depression: impact on caregivers' daily life, distress, and help

seeking. J Affect Disord 81(3):211–222. https://doi.org/10.1016/S0165-0327(03)00168-X

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. https://doi.org/10.48550/ARXIV.1706.03762

Williams A, Nangia N, Bowman S (2018) A broad-coverage challenge corpus for sentence understanding through inference. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, pp. 1112–1122. http://aclweb.org/anthology/N18-1101

Wiles N, Thomas L, Abel A, Barnes M, Carroll F, Ridgway N, Sherlock S, Turner N, Button K, Odondi L, Metcalfe C, Owen-Smith A, Campbell J, Garland A, Hollinghurst S, Jerrom B, Kessler D, Kuyken W, Morrison J, Lewis G (2014) Clinical effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: the cobalt randomised controlled trial. Health Technology Assessment (Winchester, England) 18:1–168. https://doi.org/10.3310/hta18310

Yin W, Hay J, Roth D (2019) Benchmarking zero-shot text classification: datasets, evaluation and entailment approach