Journal of Affective Disorders xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

A retrospective analysis of ketamine intravenous therapy for depression in real-world care settings

L. Alison McInnes [a,*], Jimmy J. Qian [a,b], Rishab S. Gargeya [a], Charles DeBattista [c], Boris D. Heifets [c,d]

[a] Osmind, San Francisco, CA, United States
[b] Stanford University School of Medicine, Stanford, CA, United States
[c] Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States
[d] Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, United States

ARTICLE INFO

Keywords:
Ketamine
Intravenous
Induction
Maintenance
Major depression
Real-world

ABSTRACT

Background: Outcomes of ketamine intravenous therapy (KIT) for depression in real-world care settings have been minimally evaluated. We set out to quantify treatment response to KIT in a large sample of patients from community-based practices.

Methods: We retrospectively analyzed 9016 depression patients who received KIT between 2016 and 2020 at one of 178 community practices across the United States. Depression symptoms were evaluated using the Patient Health Questionnaire-9 (PHQ-9). The induction phase of KIT was defined to be a series of 4–8 infusions administered over 7 to 28 days.

Results: Among the 537 patients who underwent induction and had sufficient data, 53.6% of patients showed a response ($\geq$ 50% reduction in PHQ-9 score) at 14–31 days post-induction and 28.9% remitted (PHQ-9 score drop to $<$ 5). The effect size was $d = 1.5$. Among patients with baseline suicidal ideation (SI), 73.0% exhibited a reduction in SI. A subset (8.4%) of patients experienced an increase in depressive symptoms after induction while 6.0% of patients reported increased SI. The response rate was uniform across 4 levels of baseline depression severity. However, more severe illness was weakly correlated with a greater drop in scores while remission status was weakly inversely correlated with depression severity. Kaplan-Meier analyses showed that a patient who responds to KIT induction has approximately 80% probability of sustaining response at 4 weeks and approximately 60% probability at 8 weeks, even without maintenance infusions.

Conclusion: KIT can elicit a robust antidepressant response in community clinics; however, a small percentage of patients worsened.

1. Introduction

Ketamine intravenous therapy (KIT) is a rapid and effective treatment for depression (McIntyre et al., 2020; Kryst et al., 2020). However, most efficacy data reflects responses to single infusions of ketamine administered to patients at academic medical centers. In this context, single doses of KIT produce a rapid decrease in depressive symptoms with positive effects peaking at 24 h. Several randomized studies have found that, within the first 3 days of a single ketamine infusion, 50–70% of patients experience a therapeutic response (i.e. > 50% reduction in symptoms on a standardized rating scale) (Zarate et al., 2006; Murrough et al., 2013a; Murrough et al., 2013bFava et al., 2018), but roughly 90% of all patients relapse within 2 weeks (Kryst et al., 2020; Newport et al., 2015).

Despite a lack of conclusive long-term data, ketamine clinics have opened across the United States offering a variety of infusion regimens. A recent analysis of 85 real-world outpatients treated at Massachusetts General Hospital over the course of 13 months suggested that repeated KIT was associated with clinically significant improvement in approximately 20% of patients (Sakurai et al., 2020). However, detailed time course data using standardized clinical instruments is still largely lacking. The retrospective analysis by Sakurai et al. (2020), along with several prospective randomized trials performed at academic centers (Murrough et al., 2013b; Singh et al., 2016; Wilkinson et al., 2018;

* Corresponding author.
E-mail address: alison@osmind.org (L.A. McInnes).

https://doi.org/10.1016/j.jad.2021.12.097
Received 19 August 2021; Received in revised form 12 December 2021; Accepted 20 December 2021

Phillips et al., 2019; Aust et al., 2019; Shiroma et al., 2020), employ similar long term ketamine treatment regimens. Typically, a patient would receive a series of closely spaced (4–8) infusions over a two-week period, known as the induction, followed by maintenance KIT at variable intervals. Prospective studies evaluating the efficacy of a ketamine induction regimen have found antidepressant response and remission rates, measured within days of the final infusion, to be similar to those observed after single ketamine infusions (Murrough et al., 2013b; Singh et al., 2016; Aust et al., 2019; Shiroma et al., 2020), though Phillips et al. (2019) found that repeated infusions may offer enhanced therapeutic benefit. Additionally, limited data suggest that the multiple-infusion ketamine induction results in an augmented durability of antidepressant response (Kryst et al., 2020; Singh et al., 2016; Aan Het Rot et al., 2010; Murrough et al., 2013b; Shiroma et al., 2020). Previous studies also support the safety and efficacy of a post-induction maintenance strategy notwithstanding relatively small samples (Voort et al., 2016; Archer et al., 2018; Wilkinson et al., 2018; Philips et al., 2019). However, it is unclear how patients enrolled in prospective, randomized trials compare to those seeking care in private practice. As the majority of patients receiving KIT are treated in community practices, it is crucial to assess KIT outcomes in these settings.

Here, we report outcomes on 9016 de-identified real-world outpatients with symptoms of depression who received KIT between 2016 and 2020 at one of 178 independent community ketamine practices across the United States. Treatment providers in participating clinics tracked mental health outcomes using a measurement-based care (online platform and the Patient Health Questionnaire-9 (PHQ-9). We determined that most clinicians in participating private practices treated patients with KIT regimens similar to those previously published, including an initial induction comprising between 4 and 8 ketamine infusions over the course of 2 to 4 weeks, and subsequent variably-spaced maintenance infusions. We present an analysis of outcomes using the PHQ-9 after KIT induction, as well as an estimate of the durability of response in this real-world population.

2. Methods

2.1. Sample

We analyzed a de-identified dataset of 9016 patients who received KIT for depression at one of 178 private practice community clinics between January 1, 2016 and December 30, 2020. The clinics in this study were chosen because they used a measurement-based care software platform with their patients. The primary measure utilized was the PHQ-9, which has demonstrated internal consistency and test-retest reliability (Kroenke et al., 2001; Löwe et al., 2004). Providers used a web-based interface to schedule the delivery of patient-reported measures. Patients were asked to complete a PHQ-9 electronically every 14 days (the recall period for the instrument). At designated days and times, the system sent text messages to patients' cell phones with reminders to complete the PHQ-9; these messages contained a link to a secure online portal wherein responses were logged. Providers could view scores on their portal as well as record ketamine infusions. The following information was collected from patient records: PHQ-9 responses (overall score as well as responses on individual line items), treatment dates, treatment types (induction or maintenance infusion), treatment notes, patient weight, treatment doses, and infusion durations. Demographic information, medication history, psychiatric history, and medical history were not available for this study. This retrospective analysis was approved by the Stanford University Institutional Review Board.

2.2. Clinical procedures

Located throughout 40 states, the 178 independent community practices that adopted the measurement-based care software used their own enrollment criteria and clinical protocols for KIT. There was variability in PHQ-9 administration and response across clinics and patients. The measurement-based care software allowed coding of treatment doses, infusion durations, treatment type (each infusion was labeled as an induction or maintenance infusion), and additional notes including use of adjunctive medications. The starting ketamine dose was usually 0.5 mg/kg infused over 40 min, although this dose varied and was generally increased in subsequent infusions. Though data regarding adjunctive medications used were not uniformly available for our analysis, providers' notes indicate that the most common adjunctive medications were given for hypertension, nausea, and anxiety. There was variability in the number of infusions patients received in both induction and maintenance phases of KIT. The price of a single infusion ranged from $300 to $690, with varying pricing structures including discounts for paying for a series of induction infusions at once. Patients were generally expected to pay out-of-pocket as KIT for depression is at best only partially covered by insurance at the time of this writing.

2.3. Outcomes and variables

The PHQ-9 was used as the primary measure of depression symptoms. Following convention, response to KIT was defined as $\geq$50% decrease in total PHQ-9 score from pretreatment status. Remission was defined as the PHQ-9 score decreasing to < 5 (Coley et al., 2020). The presence of suicidal ideation (SI) was defined as a score greater than or equal to 1 on line item 9 of the PHQ-9. Following the validated score cutoffs for the PHQ-9, we defined scores of 0–9 as none or mild depression, 10–14 as moderate depression, 15–19 as moderately severe depression, and 20–27 as severe depression (Kroenke et al., 2001). For the purposes of this study, we define induction to be a series of 4–8 infusions administered over the course of 7 to 28 days (with the additional stipulation that each infusion be labeled as an induction infusion in the software). The lower limit of 4 infusions for an induction is derived from Singh et al. (2016) and Wilkinson et al. (2018). The upper limit of 8 infusions is supported by the experience of community providers and by the convention used in intranasal esketamine induction during phase III studies (Daly et al., 2019; Popova et al., 2019). The baseline PHQ-9 was required to be within one month prior to induction. The post-induction PHQ-9 was required to be reported 14–31 days after the final induction infusion and prior to any maintenance infusion. When patients dropped out prior to completion of induction, we were able to use their last reported PHQ-9 provided it was at least two weeks after their final infusion.

2.4. Statistical analysis

The Chi-Square Test of Independence was used to determine whether there was a relationship between categorical variables (e.g. response rate). Effect size was estimated using Cohen's *d* statistic (Cohen, 1992). The point biserial correlation test was used to measure the correlation between binary outcomes (e.g. worsening, remitting) and baseline severity (Kornbrot, 2014). A Kaplan-Meier curve was generated to assess the durability of response after induction and before the initiation of maintenance infusions (Kaplan and Meier, 1958). A p-value of < 0.05 was considered significant. Analyses were performed using NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), and R (R Core Team, 2017). Data was plotted using Matplotlib (Caswell et al., 2021) and R. CI refers to confidence interval and SD refers to standard deviation.

3. Results

3.1. Subject characteristics

From our dataset of 9016 patients, we were able to analyze outcomes of KIT induction in a cohort of 537 cohort patients (Fig. 1). This cohort represents patients who met the following criteria: (1) received KIT induction consisting of 4 to 8 infusions within a 7–28 day period for KIT

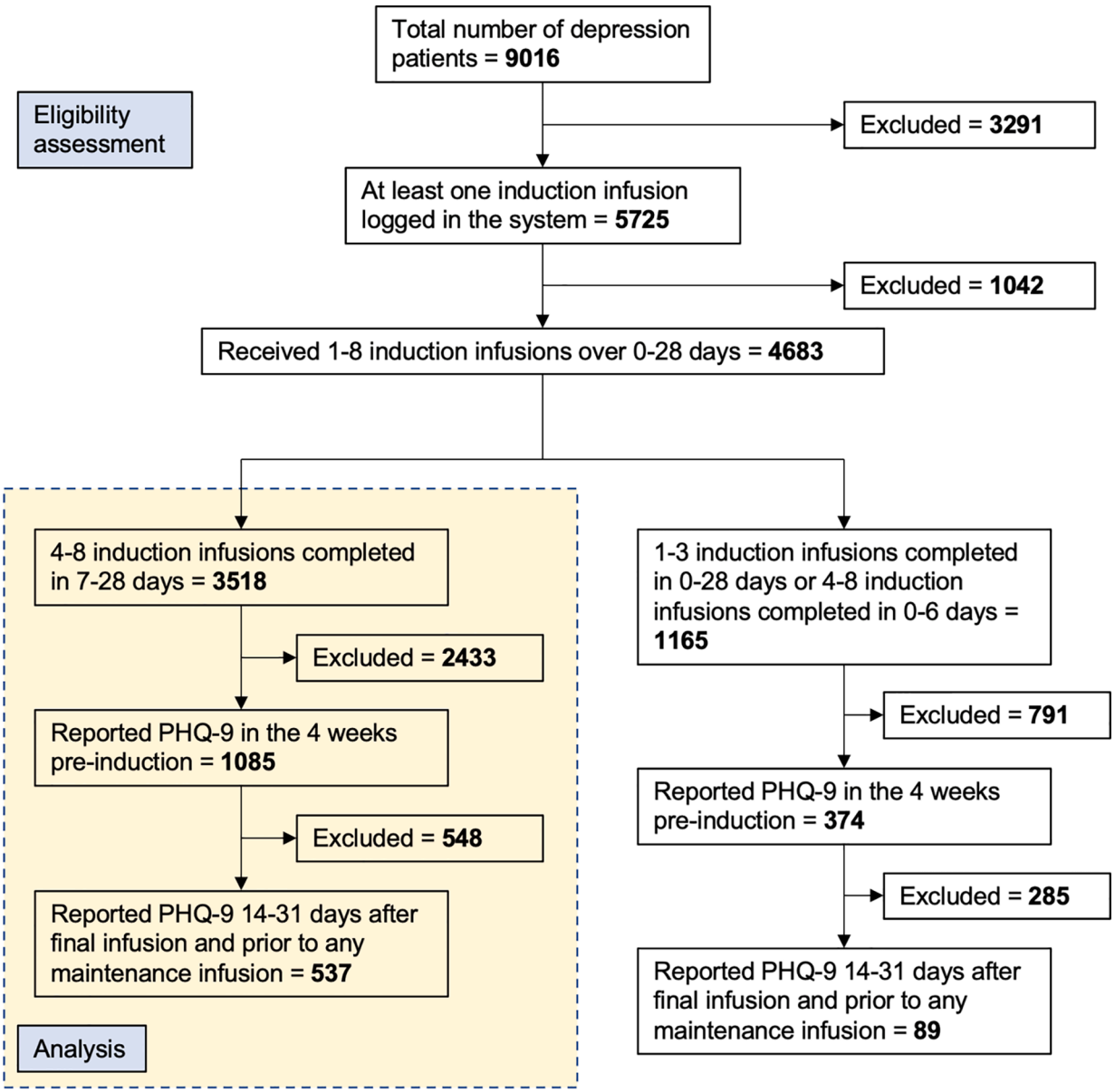

Fig. 1. Sample size and screening criteria.

induction, congruent with previously published regimens as cited above; (2) had baseline PHQ-9 data collected within one month of the initial infusion; and (3) had PHQ-9 data collected between 14–31 days after induction and prior to receiving maintenance KIT. Of the 3518 patients with 4–8 induction infusions completed within 7–28 days, 537 patients completed a PHQ-9 both before and after the induction series. Patients most commonly received 6 induction infusions in private practice settings, although there was substantial variability including a subset who did not complete induction (Supplementary Fig. 1). Supplementary Fig. 2 shows the distribution of induction length for all patients receiving 4–8 infusions. Notably, 56.9% of patients complete their induction within 14 days, consistent with published induction schemata.

We evaluated the extent to which the 537 patients included in our analysis may have differed from the overall population by comparing the distribution of baseline PHQ-9 scores in these respective groups. We tested two potential sources of selection bias leading to exclusion of a patient from analysis: dropout prior to the fourth infusion, or incomplete PHQ-9 responses.

Patients may have dropped out before the fourth infusion due to lack of antidepressant response, worsening of their depression, or inability to tolerate ketamine. To address this possibility, we compared all available baseline PHQ-9 scores for individuals who failed to complete the induction to those whom we included in our main analysis. For the 374 available baseline scores for patients in this group, the mean baseline PHQ-9 was 15.9 (SD = 6.4, 95% CI = 15.3–16.6) and the median was 16.5 (Supplementary Fig. 3). The $n = 537$ induction cohort had a mean baseline PHQ-9 score of 18.1 (SD = 5.3, 95% CI = 17.6–18.5). While the confidence intervals do not overlap, the mean and median baseline PHQ-9 scores for the $n = 374$ cohort falls within the same validated moderate-severe illness category (PHQ-9 = 15–19) as the mean and median baseline scores of all patients completing the induction (Fig. 2A). We additionally examined a subset ($n = 89$) of the 374 patients who dropped out before infusion 4 for whom we had an available PHQ-9 14–31 days after ceasing treatment early. The mean baseline PHQ-9 score in this cohort was 16.0 (SD = 5.9, 95% CI = 14.8–17.2) and the mean score 14–31 days after their final infusion was 10.3 (SD = 6.8, 95% CI = 8.9–11.7). For the $n = 537$ cohort, the mean post-induction score was 9.4 (SD = 6.5, 95% CI = 8.9–10.0). Here, the mean baseline scores also fall within the same validated moderate-severe illness category as the $n = 537$ population (Fig. 2), while the post-treatment scores have

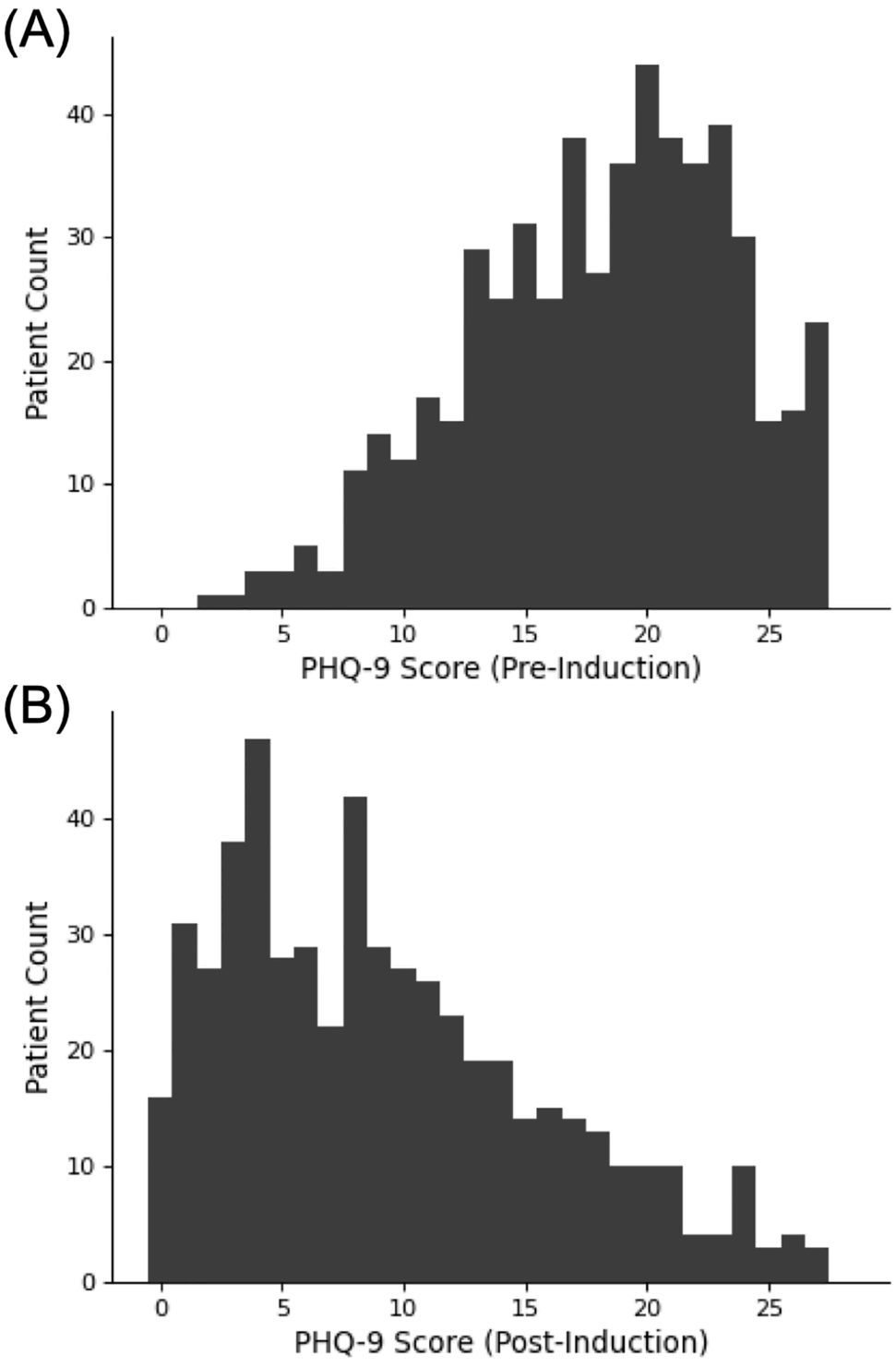

Fig. 2. PHQ-9 scores before and after induction for the $n = 537$ cohort. A) The distribution of baseline PHQ-9 scores before induction. These baseline scores had to be reported within 4 weeks of the first induction infusion. The mean and SD baseline PHQ-9 score was 18.1 Â ± 5.3. B) The distribution of baseline PHQ-9 scores after induction. The mean and SD post-induction PHQ-9 score was 9.4 Â ± 6.5. Scores were reported 14–31 days after the final induction infusion. If multiple scores were reported within either the pre- or post-induction interval by the same patient, the mean was taken.

overlapping confidence intervals and a difference in means of less than 1. Taken together, these data show that while baseline scores for patients who dropped out are statistically different from those of the $n = 537$ population, the difference is not clinically meaningful.

A second source of potential bias is that patients who did not diligently fill out surveys could not be represented in our analysis of treatment response calculation because of missing data. To address the possibility that these individuals might have influenced the response rate to produce a different outcome had they been retained in the analysis, we quantified the mean PHQ-9 score 14–31 days after induction for all 381 patients (from within the $n = 537$ cohort) who received 6 induction infusions, segmented by the number of PHQ-9 questionnaires reported by the patients (Supplementary Fig. 4). As there is no association between post-induction PHQ-9 score and the number of questionnaire responses ($\chi^2(26) = 253.0, p > 0.99$) we conclude that patients excluded for absent PHQ-9 data were not likely to systematically differ from those patients who regularly provided PHQ-9 data.

3.2. Depression outcomes for KIT induction

All post-induction outcomes were quantified 14–31 days after the final induction infusion. The $n = 537$ induction cohort had a mean

baseline PHQ-9 score of 18.1 (SD = 5.3, 95% CI = 17.6–18.5) and a mean post-induction score of 9.4 (SD = 6.5, 95% CI = 8.9–10.0) (Fig. 2). Fig. 3 shows the distribution of the drop in PHQ-9 scores. The mean decrease in raw score was 8.7 (SD = 6.6, 95% CI = 8.1–9.2, median = 8.5). The response rate to KIT induction was 53.6% (the median reduction in PHQ-9 score was 52.0%). KIT induction was associated with a Cohen's *d* effect size of 1.5. The remission rate was 28.9% (*n* = 155/537).

A subset (*n* = 45/537; 8.4%) of patients experienced an increase in PHQ-9 score during induction (Supplementary Fig. 5 shows their baseline PHQ-9 score distribution). The likelihood of worsening showed a negligible negative correlation with baseline PHQ-9 severity (point biserial correlation *r* = -0.09, *p* = 0.03) and this relationship was not clinically meaningful. Supplementary Fig. 6 shows the PHQ-9 score before and after induction for all 45 patients with worsening symptoms, including 9 patients who experienced a PHQ-9 score increase of at least 5 points.

We next examined the relationship between baseline depression severity and response to KIT induction. Fig. 4 shows the drop in PHQ-9 score as a function of baseline score. The Pearson's correlation coefficient between the drop in PHQ-9 scores and baseline PHQ-9 score was 0.44 (*p* < 10^-25) and the Spearman's correlation coefficient was 0.42 (*p*

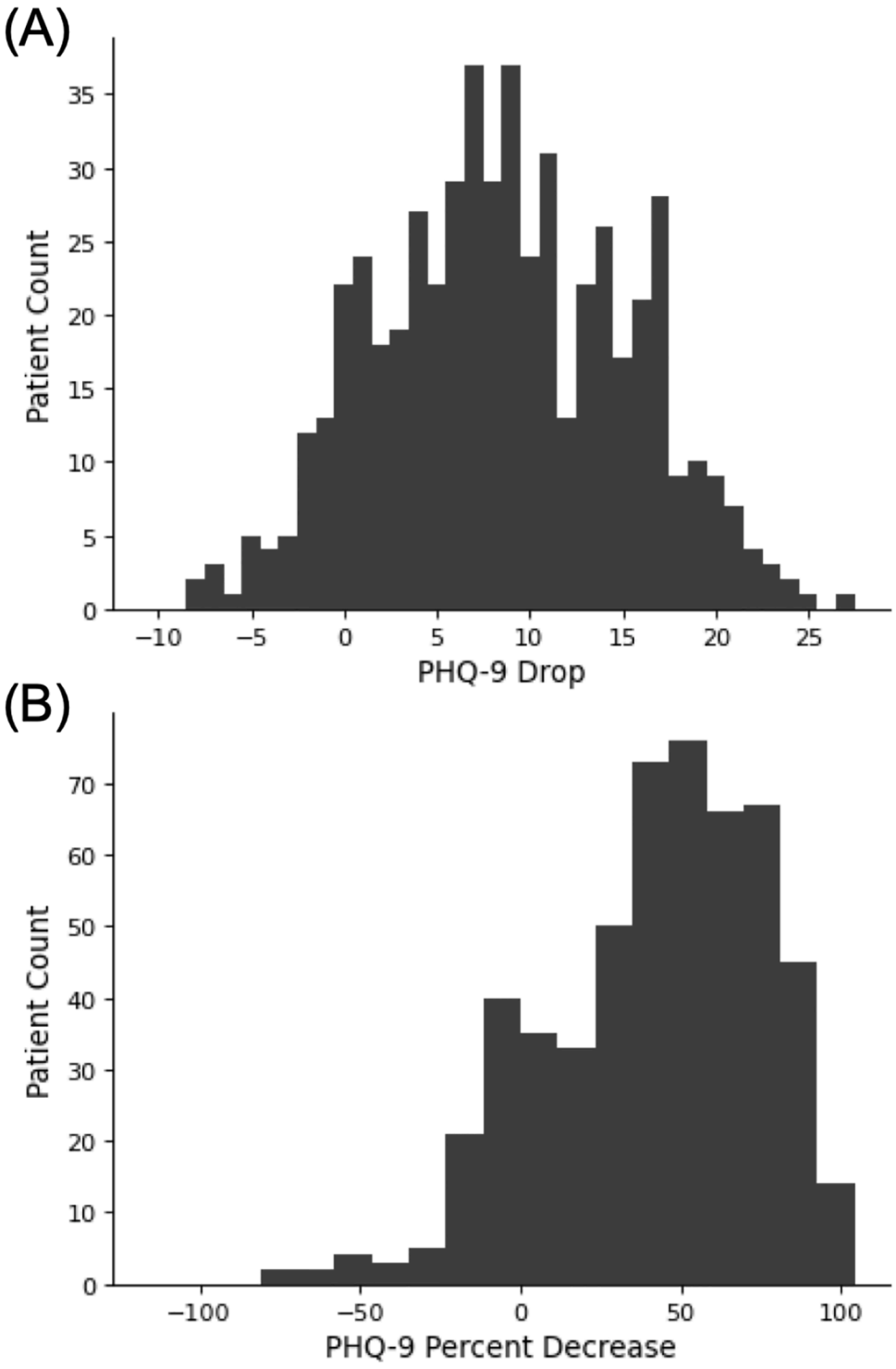

Fig. 3. Change in PHQ-9 scores for KIT induction for the *n* = 537 cohort. Data to the right of â€œ0â€ indicates an improvement in depressive symptoms. A) The distribution of the drop in PHQ-9 raw scores. The x-axis is the decrease in score. The mean and SD raw drop in score was 8.7 Â ± 6.6 (median of 8.5). B) The distribution of the percent decrease in PHQ-9 scores. The x-axis is percent decrease. The response rate was 53.6%. The mean percent drop was 47.1% and the median percent drop was 52.0%. 8.4% of patients had a PHQ-9 score that increased.

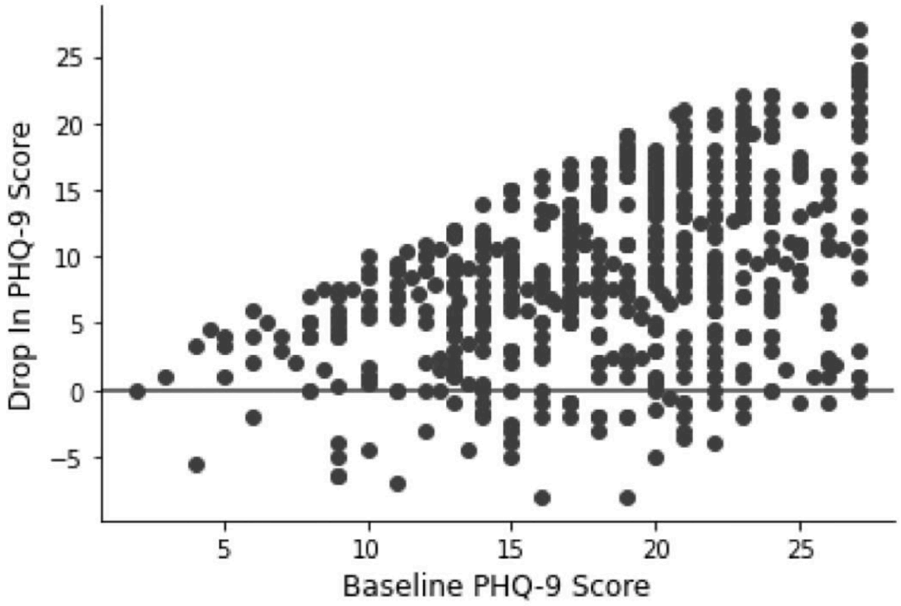

Fig. 4. The drop in PHQ-9 score as a function of baseline score. A negative drop represents worsened symptoms. Data points above the horizontal line drawn at â€œ0â€ indicate an improvement in depression severity. The Pearson's correlation coefficient between the drop in PHQ-9 scores and baseline PHQ-9 score was 0.44 ($p <$ 10^-25). The Spearman's correlation coefficient was 0.42 ($p <$ 10^-23).

< 10^-23). The response rates were 56.1%, 54.1%, 52.2%, and 53.9%, respectively, for the none/mild, moderate, moderately severe, and severe baseline levels of depression (n = 41, 98, 157, 241, respectively). The response rate did not differ based on the level of baseline depression ($\chi^2(8) = 0.24$, $p > 0.999$). These data suggest a relationship between the magnitude of reduction in PHQ-9 score and baseline symptom severity, as patients with higher baseline PHQ-9 scores would need to achieve larger reductions in score than patients with lower baseline scores to achieve responder status.

The likelihood of remission was negatively correlated with the 4 baseline levels of severity at 58.5%, 43.9%, 28.7%, and 17.8%, respectively ($\chi^2(8) = 42.6$, $p <$ 10^-5). Although there was no difference in response rate based on baseline severity, more severe illness was weakly inversely correlated with the likelihood of remission (point biserial correlation r = -0.27, $p <$ 10^-9). The baseline PHQ-9 scores for remitters are shown in Supplementary Fig. 7.

The percentages of patients per illness severity category who had a PHQ-9 score increase were 14.6%, 8.2%, 10.8%, and 5.8%, respectively. There was no correlation between baseline symptom severity and tendency to worsen ($\chi^2(8) = 5.395$, $p = 0.71$).

Supplementary Table 1 shows outcomes grouped by the number of infusions received during induction.

3.3. SI in KIT induction

Within the n = 537 cohort, 66.3% (n = 356) had baseline SI to some degree, defined as a score > 0 on PHQ9 question 9, and 73.0% of this subset (260/356) reported an improvement in SI after KIT induction. Moreover, 42.7% (152/356) of patients who had SI prior to induction exhibited no more SI at the end of induction. Among the 356 patients with baseline SI, 79.6% (86/108), 85.6% (83/97), and 60.3% (91/151) of patients with an average baseline PHQ-9 line item 9 score of 3, 2, 1, respectively, exhibited a decrease in SI. The distributions of SI at baseline and after induction are shown in Supplementary Fig. 8. Within the n = 537 cohort, 16 individuals experienced an increase in both SI and overall PHQ-9 score, while 16 individuals experienced an increase in SI but no increase in overall PHQ-9 score. In other words, 35.6% (16/45) of patients who worsened in terms of global depression symptoms also reported an increase in SI, and 6.0% of the overall n = 537 cohort reported an increase in SI.

3.4. Durability of response to induction

For patients who either responded or remitted after induction, we performed a Kaplan-Meier analysis to estimate the probability of relapse over time (Fig. 5). An event represents a patient losing response (relapse). We performed right censoring for those patients who entered maintenance or were lost to follow up (defined as failing to submit a PHQ-9 every two weeks). The y-axis represents the probability of retaining responder status and the x-axis is days after the last induction infusion. We observed that the probability of retaining responder status at 28 days after induction is 78.3% (95% CI = 73.3–83.7%) and the probability at 56 days is 59.9% (95% CI = 51.7%-69.4%). Table 1 summarizes the probability of maintaining an antidepressant response and uncertainty estimates up to 91 days post-induction without maintenance infusions.

3.5. Practice patterns for maintenance KIT

After induction, patients enter the maintenance phase of KIT. However, little is known about how often patients come in for follow-up or for how long they remain in treatment. Details of the timing and number of maintenance infusions can be found in Supplementary Tables 2–4. 52.5% of patients who completed induction entered maintenance treatment. On average, patients who completed induction and responded to treatment received 3.6 maintenance infusions (SD = 4.9, 95% CI = 3.0–4.2) while remitters received 3.7 (SD = 5.0, 95% CI = 2.9–4.5). The 3518 patients who completed KIT induction (regardless whether they reported sufficient PHQ-9 scores to be included in the n = 537 cohort) received 2.6 maintenance infusions (SD = 5.0, 95% CI = 2.5–2.8) on average.

The mean length between the final induction infusion and the first maintenance infusion for the 109 patients who completed induction, achieved remission, and entered maintenance was 84.3 days (95% CI = 66.6–102.0, median = 48) and the mean length for the 202 responders who entered maintenance was 75.1 days (95% CI = 63.1–87.2, median = 43). For all 1846 patients who completed induction and entered maintenance, the mean length was 46.3 days (95% CI = 42.2–50.3,

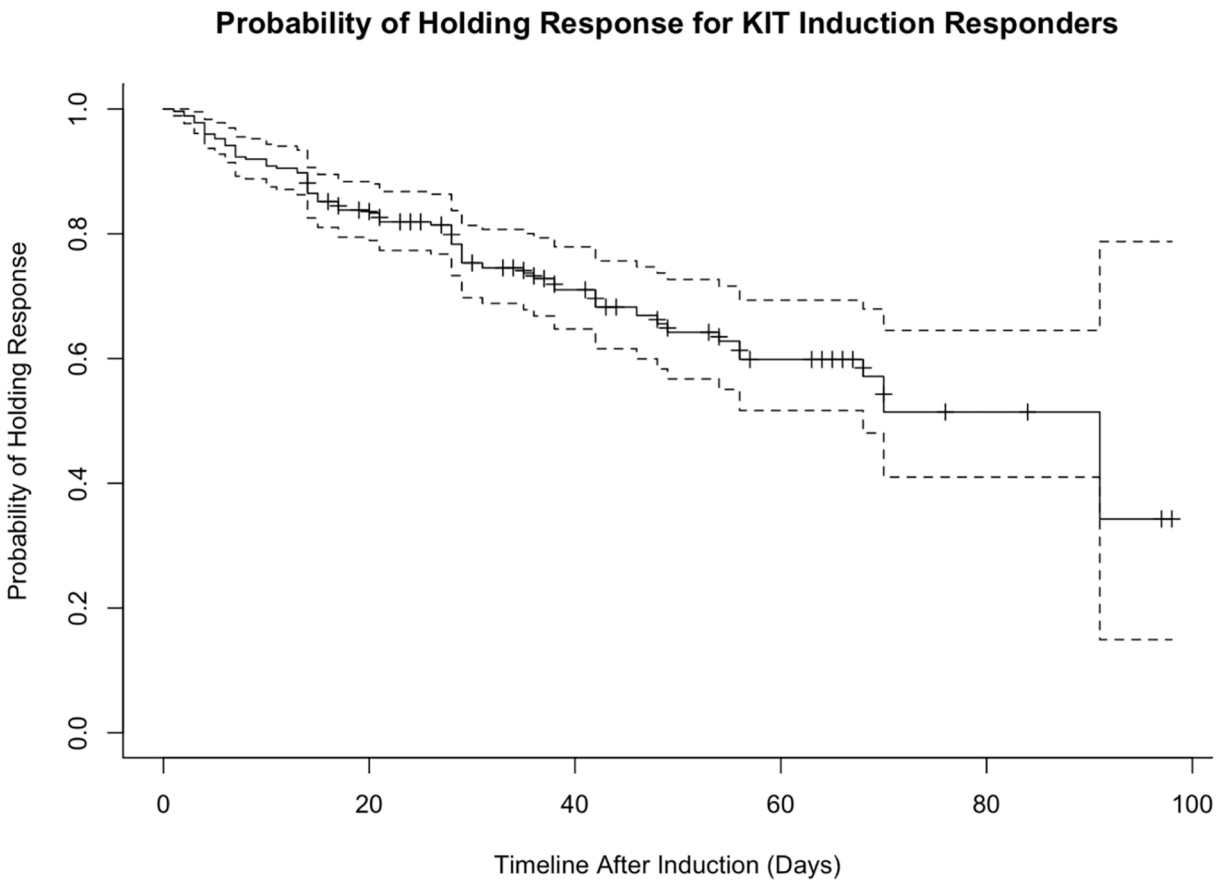

Fig. 5. Kaplan-Meier estimates of the probability that a patient who responded to induction has not lost responder status over time. Loss of response (the 'event' in the survival curve) is defined as the PHQ-9 score increasing to the point that there is no longer a â‰¥50% improvement from baseline. Patients were censored when they began receiving maintenance infusions or when they stopped reporting at least one PHQ-9 every two weeks. Vertical lines indicate censored observations. The solid line is the survival curve while dotted lines represent 95% confidence intervals on the survival curve. The x-axis is the number of days since the end of induction. Of the 288 patients who responded to induction, 274 had sufficient data for this survival analysis. Of these, 76 patients experienced a loss of response, 125 patients were lost to follow-up, and 73 patients entered maintenance treatment. Table 1 shows the numerical details of this survival analysis.

median = 28).

4. Discussion

This is the largest published analysis to date examining the real-world effectiveness of a standard KIT induction protocol for depression using data from patients treated at community clinics across the United States. From a dataset of 9016 patients, we focused on a cohort of 537 individuals who underwent a KIT induction as defined above, and for whom sufficient outcomes data before and after induction were available. We found that response to KIT induction is both robust and durable.

We observed an overall response rate of 53.6% and a remission rate of 28.9% measured at 14–31 days after the last infusion. This contrasts with a recent study of 85 community outpatients by Sakurai et al. (2020) who reported a response rate of only 18.5% measured immediately prior to the 6th infusion using the Quick Inventory of Depressive Symptomatology-Self Report scale (QIDS-SR16). Possible reasons for discrepant findings include the timing of the final QIDS-SR16, e.g. missing the 6th infusion, and the fact that they administered all their infusions between 5:30 and 8 pm while most community patients in our study received KIT during daytime business hours. It has been posited that the time of day in which KIT is administered could potentially influence antidepressant response given that ketamine has an effect on circadian rhythms (Zhuo et al., 2019; Orozco-Solis et al., 2017; Bellet et al., 2011).

While our response rates diverge from those of the Sakurai group, our data is roughly in line with that of Phillips et al. (2019) Singh et al. (2016), Aust et al. (2019) and Wilkinson et al. (2018) who observed response rates between 45 and 59% after an induction protocol. We note that most studies we cite measured induction response just after the final infusion, whereas our measure was 14–31 days after the final infusion and thus may be an underestimate of the true response rate if our sample were measured at a comparably early post-induction timepoint.

It is possible that the response rate we observed could be skewed upwards by unblinded clinicians, who want their patients to improve, and patients, who paid for treatment with hopes of feeling better. This potential bias is an intrinsic limitation of real-world analyses vis-a-vis clinical trials. On the other hand, Sakurai et al. (2020) reported a relatively low response rate from data garnered under similar conditions (patient payment for treatment, real world setting), suggesting that any possible upward skewing of response rates is unlikely to fully explain our results.

We found that response rates did not vary as a function of initial depression severity, though more severely ill patients tended to show a greater drop in PHQ-9 scores while also having a reduced likelihood of remission. As with Sakurai et al. (2020), we also found that a small subset of patients (8.4%) who completed induction worsened. There was no relationship between baseline PHQ-9 score and the likelihood of worsening and only about $^{1}/_{3}$ of patients who worsened experienced an increase in SI. Overall, most patients with SI at baseline experienced an improvement in this symptom and 43% of those with SI no longer had SI after induction; however, 6.0% of patients reported an increase in SI after induction. Reasons for worsening of symptoms could include a

Table 1
Summary of Kaplan-Meier analysis presented in Fig. 5.

Time (days after infusion)	Number of patients at risk	Number of events	Survival probability	Standard error	Lower 95% confidence interval	Upper 95% confidence interval
1	274	1	0.996	0.00364	0.989	1.000
2	273	2	0.989	0.00629	0.977	1.000
3	271	3	0.978	0.00884	0.961	0.996
4	268	5	0.960	0.01186	0.937	0.983
5	263	2	0.953	0.01284	0.928	0.978
6	261	3	0.942	0.01417	0.914	0.970
7	258	5	0.923	0.01607	0.892	0.955
8	253	1	0.920	0.01642	0.888	0.952
10	252	3	0.909	0.01740	0.875	0.944
11	249	1	0.905	0.01770	0.871	0.940
13	248	2	0.898	0.01830	0.863	0.934
14	246	9	0.865	0.02065	0.825	0.906
15	195	3	0.852	0.02171	0.810	0.895
17	187	3	0.838	0.02275	0.795	0.884
20	180	1	0.833	0.02310	0.789	0.880
21	176	3	0.819	0.02411	0.773	0.868
26	161	1	0.814	0.02450	0.767	0.863
28	159	6	0.783	0.02659	0.733	0.837
29	104	4	0.753	0.02953	0.697	0.813
31	96	1	0.745	0.03024	0.688	0.807
35	87	1	0.737	0.03109	0.678	0.800
36	85	1	0.728	0.03191	0.668	0.793
38	81	2	0.710	0.03355	0.647	0.779
42	77	3	0.682	0.03585	0.616	0.756
46	51	1	0.669	0.03756	0.599	0.747
48	50	1	0.656	0.03912	0.583	0.737
49	48	1	0.642	0.04062	0.567	0.727
54	45	1	0.628	0.04215	0.550	0.716
56	43	2	0.599	0.04496	0.517	0.694
68	22	1	0.571	0.05048	0.481	0.679
70	20	2	0.514	0.05944	0.410	0.645
91	3	1	0.343	0.14546	0.149	0.787

dysphoric anxious response to ketamine (Aust et al., 2019), complications of comorbid diagnoses including affective switching (Bhatt et al., 2021) or other complex comorbidities (Niciu et al., 2013), failure to tolerate ketamine side-effects such as dissociation or nausea, and a fear of what might happen next if a "treatment of last resort" fails. Our findings highlight that practitioners should exercise caution when treating only mildly ill individuals as KIT is not without potential adverse consequences.

In addition to the robust response and remission rates, we also observed a strong durability of response. At 4 weeks after induction, there is approximately 80% probability that responders do not lose responder status even without maintenance infusions, and at 8 weeks the probability is approximately 60% (Fig. 5). The median time to a first maintenance infusion was 48 days for remitters and 43 days for responders, versus 28 days for all 1846 patients who completed induction and entered maintenance. The sustained response to repeated infusions that we observed contrasts with the transient response to single infusions (Kryst et al., 2020; Newport et al., 2015), further supporting the utility of the KIT induction model that has become widely adopted. We again acknowledge that the expense of treatment may lead to an upward bias in duration of response in any real-world treatment dataset. Additionally, these results diverge from the intranasal esketamine model which requires weekly or biweekly treatments to maintain response after the induction (Daly et al., 2019), although intranasal administration of esketamine is not necessarily comparable to KIT using racemic ketamine. At this time, there is insufficient comparative efficacy data to draw further conclusions.

We and Sakurai et al. (2020) both note that patients entering maintenance received a relatively small number of maintenance infusions (roughly 2.6 in our broad data set and 2.3 in the latter study) with a median time interval between them of 28 days in our data set (n = 1309). Thus, most patients tend to exit treatment within 6 months although there is substantial variation. It is not clear whether patients exit treatment because they feel better or they can no longer afford the treatment. It has been suggested that response to KIT may be enhanced with psychotherapy during the maintenance phase if patients are able to make healthy new lifestyle changes (Greenway et al., 2020), and some patients may no longer feel they need maintenance KIT once those changes are implemented. Some community-based practitioners have advocated for a style of ketamine therapy called "Ketamine-Assisted Psychotherapy" (KAP) which incorporates behavioral and psychotherapeutic interventions pre-treatment, on-drug, and during the post-induction period (Dore et al., 2019). Use of ameasurement based care platform will facilitate a more rigorous evaluation of response durability with these differing models of ketamine therapy used in community practices.

5. Limitations

Our analysis has some major limitations. We report outcomes on a fraction of the total sample due to missing mood survey data. Of the 3518 patients identified who completed KIT induction, only 15.3% had PHQ-9 data available both pre- and post-treatment. This particular measurement-based care software did not collect demographic information, past medical history, or substance use data. We lack data on the chronicity of the episode for which the patient is being treated, number of failed traditional antidepressants, previous electroconvulsive therapy or transcranial magnetic stimulation treatment, and current medications including those used to manage side-effects during KIT. We also do not know who was receiving psychotherapy, which may make ketamine and other psychedelic agents more effective. While it is assumed that patients undergoing induction treatments are ketamine naive, that was not documented explicitly. Dosage information was not available for many treatments, though we believe it is a fair assumption that KIT dosages started at 0.5 mg/kg and increased throughout the induction. There is an economic and racial selection bias for patients who are able to access KIT. There are also inherent limitations with retrospective analyses compared to randomized controlled trials, though our real-world study

does present many important advantages in elucidating outcomes of KIT in diverse naturalistic community settings.

Several of the limitations acknowledged here can be addressed by future replication studies that incorporate more detailed data collected in Osmind's electronic health record (EHR) platform. The present analysis utilized a dataset from a separate measurement-based care platform, which was not a full EHR and thus did not capture various data variables that would enhance our understanding of KIT in real-world practice settings. The Osmind EHR captures demographic data and clinical history, including number and duration of episodes of depression and a full medication history, and also automates outcomes tracking with a patient facing mobile phone application to facilitate measurement based care by clinicians. This richer data set can form the basis of predictive models that directly inform clinical care. For example, to what degree does treatment refractoriness affect response and relapse? How do medications, taken daily or as adjuncts to KIT, modify response? Can a simplified set of PHQ9 items predict early response versus treatment failure? More broadly, point of care data collection and personal sensing streams (e.g. actigraphy, or "wearables") can establish links between self report measures and functional health outcomes. Finally, as KIT practices grow in popularity with substantial variability in care models, it is imperative to define long-term patient outcomes, data which is at present completely absent in the literature.

6. Conclusions

We observed a robust and durable antidepressant response to KIT 14–31 days post-induction in 53.6% of those that completed the induction (effect size was $d = 1.5$). Responders to KIT induction have approximately 80% probability of sustaining response at 4 weeks and approximately 60% probability at 8 weeks, even without maintenance infusions. KIT was also associated with remission for 28.9% of patients completing induction. 42.7% of patients who had SI at baseline no longer experienced SI after induction. 52.5% of patients completing induction elected to continue into maintenance treatment with the average patient receiving 2–3 maintenance infusions. Patients with the most severe illness respond equally well as those with less severe illness, but have lower rates of remission. A small percentage of people worsen with KIT and this should be an area of future research.

Role of funding

Research reported in this publication was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR003142. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Previous presentation

None.

CRediT authorship contribution statement

L. Alison McInnes: Visualization, Data curation, Writing – original draft. **Jimmy J. Qian:** Visualization, Data curation, Writing – original draft. **Rishab S. Gargeya:** Formal analysis, Data curation, Writing – original draft. **Charles DeBattista:** Visualization, Data curation, Writing – original draft. **Boris D. Heifets:** Visualization, Data curation, Writing – original draft.

Declaration of Competing Interest

LAM is a consultant to Clexio Biosciences and an employee of Osmind Inc. JJQ is an employee of Osmind Inc. RSG is an employee of Osmind Inc. CD is a consultant to Magnus Medical, Corcept Therapeutics, Sage Therapeutics, Pfizer, Alkermes, and Osmind; he has received research support from Sage Therapeutics, Compass Pathways, Relmada Therapeutics, and Janssen. BDH is not supported by, nor maintains any financial interest in, any commercial activity that may be associated with the topic of this article.

Acknowledgment

We thank Balasubramanian Narasimhan, PhD, Bradley Efron, PhD, Alda Pontes, and Eric C. Sun, MD, PhD, for their advice regarding statistical analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2021.12.097.

References

Aan Het Rot, M., Collins, K.A., Murrough, J.W., Perez, A.M., Reich, D.L., Charney, D.S., Mathew, S.J., 2010. Safety and efficacy of repeated-dose intravenous ketamine for treatment-resistant depression. Biol. Psychiatry 67, 139–145. https://doi.org/10.1016/j.biopsych.2009.08.038.

Archer, S., Chrenek, C., Swainson, J., 2018. Maintenance ketamine therapy for treatment-resistant depression. J. Clin. Psychopharmacol. 38, 380–384. https://doi.org/10.1097/JCP.0000000000000894.

Aust, S., Gärtner, M., Basso, L., Otte, C., Wingenfeld, K., Chae, W.R., Heuser-Collier, I., Regen, F., Cosma, N.C., van Hall, F., Grimm, S., 2019. Anxiety during ketamine infusions is associated with negative treatment responses in major depressive disorder. Eur. Neuropsychopharmacol. 29, 529–538. https://doi.org/10.1016/j.euroneuro.2019.02.005.

Bhatt, K., Yoo, J., Bridges, A., 2021. Ketamine-induced manic episode. Prim. Care Companion CNS Disord. 23, 20102811 https://doi.org/10.4088/PCC.20l02811.

Bellet, M.M., Vawter, M.P., Bunney, B.G., Bunney, W.E., Sassone-Corsi, P., 2011. Ketamine influences CLOCK: BMAL1 function leading to altered circadian gene expression. PLoS One 6, e23982. https://doi.org/10.1371/journal.pone.0023982.

Caswell, T.A., Droettboom, M., Lee, A., de Andrade, E.S., Hunter, J., Hoffmann, T., et al., 2021. matplotlib/matplotlib: REL: v3.4.0rc3 (Version v3.4.0rc3). Zenodo. 10.5281/zenodo.4595937.

Cohen, J., 1992. Statistical power analysis. Curr. Dir. Psychol. Sci. 1, 98–101. https://doi.org/10.1111/1467-8721.ep10768783.

Coley, R.Y., Boggs, J.M., Beck, A., Hartzler, A.L., Simon, G.E., 2020. Defining success in measurement-based care for depression: a comparison of common metrics. Psychiatr. Serv. 71, 312–318. https://doi.org/10.1176/appi.ps.201900295.

Daly, E.J., Trivedi, M.H., Janik, A., Li, H., Zhang, Y., Li, X., Lane, R., Lim, P., Duca, A.R., Hough, D., Thase, M.E., et al., 2019. Efficacy of esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression: a randomized clinical trial. JAMA Psychiatry 76, 893–903. https://doi.org/10.1001/jamapsychiatry.2019.1189.

Dore, J., Turnipseed, B., Dwyer, S., Turnipseed, A., Andries, J., Ascani, G., Monnette, C., Huidekoper, A., Strauss, N., Wolfson, P., 2019. Ketamine assisted psychotherapy (KAP): patient demographics, clinical data and outcomes in three large practices administering ketamine with psychotherapy. J. Psychoact. Drugs 51, 189–198. https://doi.org/10.1080/02791072.2019.1587556.

Greenway, K.T., Garel, N., Jerome, L., Feduccia, A.A., 2020. Integrating psychotherapy and psychopharmacology: psychedelic-assisted psychotherapy and other combined treatments. Expert Rev. Clin. Pharmacol. 13, 655–670. https://doi.org/10.1080/17512433.2020.1772054.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., et al., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2Ionescu.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. 53, 457–481. https://doi.org/10.1080/01621459.1958.10501452.

(eds Kornbrot, D., Balakrishnan, N., 2014. Point biserial correlation. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat06227 (edset al., 2014.

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

Kryst, J., Kawalec, P., Mitoraj, A.M., Pilc, A., Lasoń, W., Brzostek, T., 2020. Efficacy of single and repeated administration of ketamine in unipolar and bipolar depression: a meta-analysis of randomized clinical trials. Pharmacol. Rep. 72, 543–562. https://doi.org/10.1007/s43440-020-00097-z.

Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med. Care 42, 1194–1201.

McIntyre, R.S., Carvalho, I.P., Lui, L.M., Majeed, A., Masand, P.S., Gill, H., Rodrigues, N. B., Lipsitz, O., Coles, A.C., Lee, Y., Tamura, J.K., 2020. The effect of intravenous,

intranasal, and oral ketamine/esketamine in mood disorders: a meta-analysis. J. Affect. Disord. 276, 576–584. https://doi.org/10.1016/j.jad.2020.06.050.

Murrough W, J, et al., 2013a. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site randomized controlled trial. The American Journal of Psychiatry 170 (10), 1134–1142. https://doi.org/10.1176/appi.ajp.2013.13030392.

Niciu, M.J., Grunschel, B.D., Corlett, P.R., Pittenger, C., Bloch, M.H., 2013. Two cases of delayed-onset suicidal ideation, dysphoria and anxiety after ketamine infusion in patients with obsessive–compulsive disorder and a history of major depressive disorder. J. Psychopharmacol. 27, 651–654. https://doi.org/10.1177/0269881113486718.

Murrough, J.W., Perez, A.M., Pillemer, S., Stern, J., Parides, M.K., Aan Het Rot, M., Collins, K.A., Mathew, S.J., Charney, D.S., Iosifescu, D.V., 2013b. Rapid and longer-term antidepressant effects of repeated ketamine infusions in treatment-resistant major depression. Biol. Psychiatry 74, 250–256. https://doi.org/10.1016/j.biopsych.2012.06.022.

Newport, D.J., Carpenter, L.L., McDonald, W.M., Potash, J.B., Tohen, M., Nemeroff, C.B., APA Council of Research Task Force on Novel Biomarkers and Treatments, 2015. Ketamine and other NMDA antagonists: early clinical trials and possible mechanisms in depression. Am. J. Psychiatry 172, 950–966. https://doi.org/10.1176/appi.ajp.2015.15040465.

Orozco-Solis, R., Montellier, E., Aguilar-Arnal, L., Sato, S., Vawter, M.P., Bunney, B.G., Bunney, W.E., Sassone-Corsi, P., 2017. A circadian genomic signature common to ketamine and sleep deprivation in the anterior cingulate cortex. Biol. Psychiatry 82, 351–360. https://doi.org/10.1016/j.biopsych.2017.02.1176.

Phillips, J.L., Norris, S., Talbot, J., Birmingham, M., Hatchard, T., Ortiz, A., Owoeye, O., Batten, L.A., Blier, P., 2019. Single, repeated, and maintenance ketamine infusions for treatment-resistant depression: a randomized controlled trial. Am. J. Psychiatry 176, 401–409. https://doi.org/10.1176/appi.ajp.2018.18070834.

Popova, V., Daly, E.J., Trivedi, M., Cooper, K., Lane, R., Lim, P., Mazzucco, C., Hough, D., Thase, M.E., Shelton, R.C., Molero, P., et al., 2019. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. Am. J. Psychiatry 176, 428–438. https://doi.org/10.1176/appi.ajp.2019.19020172.

R Core Team, 2017. A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Sakurai, H., Jain, F., Foster, S., Pedrelli, P., Mischoulon, D., Fava, M., Cusin, C., 2020. Long-term outcome in outpatients with depression treated with acute and maintenance intravenous ketamine: a retrospective chart review. J. Affect. Disord. 276, 660–666. https://doi.org/10.1016/j.jad.2020.07.089.

Shiroma, P.R., Thuras, P., Wels, J., Albott, C.S., Erbes, C., Tye, S., Lim, K.O., 2020. A randomized, double-blind, active placebo-controlled study of efficacy, safety, and durability of repeated vs single subanesthetic ketamine for treatment-resistant depression. Transl. Psychiatry 10, 206. https://doi.org/10.1038/s41398-020-00897-0.

Singh, J.B., Fedgchin, M., Daly, E.J., De Boer, P., Cooper, K., Lim, P., Pinter, C., Murrough, J.W., Sanacora, G., Shelton, R.C., Kurian, B., Winokur, A., Fava, M., Manji, H., Drevets, W.C., Van Nueten, L., 2016. A double-blind, randomized, placebo-controlled, dose-frequency study of intravenous ketamine in patients with treatment-resistant depression. Am. J. Psychiatry 173, 816–826. https://doi.org/10.1176/appi.ajp.2016.16010037.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Voort, J.L.V., Morgan, R.J., Kung, S., Rasmussen, K.G., Rico, J., Palmer, B.A., Schak, K. M., Tye, S.J., Ritter, M.J., Frye, M.A., Bobo, W.V., 2016. Continuation phase intravenous ketamine in adults with treatment-resistant depression. J. Affect. Disord 206, 300–304. https://doi.org/10.1016/j.jad.2016.09.008.

Wilkinson, S.T., Katz, R.B., Toprak, M., Webler, R., Ostroff, R.B., Sanacora, G., 2018. Acute and longer-term outcomes using ketamine as a clinical treatment at the Yale psychiatric hospital. J. Clin. Psychiatry 79. https://doi.org/10.4088/JCP.17m11731.

Zhuo, C., Tian, H., Li, G., Chen, M., Jiang, D., Lin, X., Xu, Y., Wang, W., 2019. Effects of ketamine on circadian rhythm and synaptic homeostasis in patients with treatment-resistant depression: a protocol for mechanistic studies of its rapid and sustained antidepressant actions in humans. Brain Behav. 9, e01423. https://doi.org/10.1002/brb3.1423.