

Method

Statistical analyses

All statistical analyses were performed using STATA 14.2 (College Station, TX). We chose to use a mixed-effects model, as we expected heterogeneity between studies based on slightly different protocols and inclusion/exclusion criteria. To examine the heterogeneity of studies, we calculated an F-statistic as the ratio of the residual sum of squares comparing models that did and did not nest subjects within studies. Because of small numbers, the race variable was collapsed as white versus non-white, and the diagnosis variable was collapsed as MDD (unipolar) versus other diagnoses.

Combining data from different scales

In order to combine composite depression scores across studies where different rating scales were used, we used a conversion table to convert HAM-D scores to MADRS scores, as previously described (1). Individual suicidal ideation items were combined by multiplying the HAM-D suicidal ideation item scores by 1.5 to convert them to MADRS scores. We recognize that there is no precedent for combining individual suicidal ideation item scores for the MADRS (scale of 0-6) and HAM-D (scale of 0-4) in this manner; thus, we also present a sensitivity analysis where the results for the MADRS and HAM-D suicide items were examined separately. Because the QIDS-SR and BDI suicide items have the same range, we performed a combined analysis with no conversion factor.

Approach to missing data

The general linear mixed model was robust to missing data and hence allowed us to use all available subjects. Three studies did not collect data at all time points of interest (baseline, and Days 1, 2, 3, and 7). Specifically, Berman and colleagues (2) did not collect data at Day 7, and Hu and colleagues (3) and Sos and colleagues (4) did not collect data on Day 2. Overall, missing data from these sources accounted for 9.5% of all data. These data were considered missing not at random. For these missing data, imputation was done by cross-sectional average stratified by treatment. We felt this was justified as there was no evidence of significant heterogeneity between studies ($F_{796}=1.00$, $p=0.500$) and because—after stratifying for treatment—there were no differences at any time points between these three studies that did not collect data on certain days and other studies. After performing this imputation, 5.3% of data remained missing. Because ‘missingness’ was not related to treatment assignment or clinical response, we felt these data were missing completely at random. No further imputation was performed for the remaining 5.3% of the missing data.

Dichotomous Outcomes

Dichotomous outcomes (being free of suicidal ideation) were analyzed in two separate models (self-reported and clinician-administered outcome measures). Subjects who had a 0 or 1 on the MADRS item 10 were considered free of suicidal ideation; subjects who had a 0 on the HAM-D item 3, QIDS-SR item 12, and/or BDI item 9 were also considered free of suicidal ideation. The proportion of subjects free of suicidal ideation at each time point between groups for each assessment was compared using two-sample proportional tests.

Supplemental References

1. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol.* 2006;16:601-611.
2. Berman RM, Cappiello A, Anand A, Oren DA, Heninger GR, Charney DS, Krystal JH. Antidepressant effects of ketamine in depressed patients. *Biological psychiatry.* 2000;47:351-354.
3. Hu YD, Xiang YT, Fang JX, Zu S, Sha S, Shi H, et al. Single i.v. ketamine augmentation of newly initiated escitalopram for major depression: results from a randomized, placebo-controlled 4-week study. *Psychol Med.* 2015:1-13.
4. Sos P, Klirova M, Novak T, Kohutova B, Horacek J, Palenicek T. Relationship of ketamine's antidepressant and psychotomimetic effects in unipolar depression. *Neuro endocrinology letters.* 2013;34:287-293.

TABLE S1. Clinical Ratings Used Across Trials

Reference	Total N	Included N	Included Clinical Ratings
Berman et al 2000 (10)	8	5	HAMD, BDI
Valentine et al 2011 (25)	11	4	HAMD, BDI
Murrough et al 2013 (14)	73	35	MADRS, HAMD, QIDS-SR
Sos et al 2013 (24)	27	9	MADRS, BDI
Feder et al 2014 (22)	41	5	MADRS, QIDS-SR
Ballard et al 2014 (15)	87	59	MADRS, HAMD, BDI
Hu et al 2015 (23)	27	26	MADRS, QIDS-SR
Murrough et al 2015 (18)	24	24	MADRS, QIDS-SR
Total	298	167	

Abbreviations: HAM-D: Hamilton Depression Rating Scale; MADRS: Montgomery Asberg Depression Rating Scale; BDI: Beck Depression Inventory; QIDS: Quick Inventory of Depressive Symptomatology-Self Report

TABLE S2. Demographics of Patients Included in Meta-Analysis

Reference	Included N	Age		Female		White		MDD Diagnosis	
		Mean	SD	N	%	N	%	N	%
Berman et al 2000 (10)	5	36.6	8.9	2	40	4	80	5	100
Valentine et al 2011 (25)	4	50.6	5.4	2	50	2	50	4	100
Sos et al 2013 (24)	9	42.4	14.3	4	44	9	100	9	100
Murrough et al 2013 (14)	35	43.9	12.3	16	46	28	80	35	100
Feder et al 2014 (22)	5	33.4	11.3	2	40	1	20	0	0
Ballard et al 2014 (15)	59	43.8	11.0	34	58	46	78	33	56
Hu et al 2015 (23)	26	38.9	12.9	17	65	0	0	26	100
Murrough et al 2015 (18)	24	42.4	13.3	16	67	21	88	16	67
Total	167	42.4	12.1	93	57	111	66	128	77