Research paper

Parsing the heterogeneity of depression: An exploratory factor analysis across commonly used depression rating scales

Elizabeth D. Ballard[a,*], Julia S. Yarrington[a], Cristan A. Farmer[a], Marc S. Lener[a], Bashkim Kadriu[a], Níall Lally[a,b,c], Deonte Williams[a], Rodrigo Machado-Vieira[a,d], Mark J. Niciu[a], Lawrence Park[a], Carlos A. Zarate Jr[a]

[a] Experimental Therapeutics & Pathophysiology Branch, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA
[b] Institute of Cognitive Neuroscience, University College London, London WC1N 3AZ, UK
[c] Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK
[d] Dept. of Psychiatry and Behavioral Sciences, University of Texas Health Sciences Center, Houston, TX 77021, USA

ABSTRACT

Background: Due to the heterogeneity of depressive symptoms—which can include depressed mood, anhedonia, negative cognitive biases, and altered activity levels—researchers often use a combination of depression rating scales to assess symptoms. This study sought to identify unidimensional constructs measured across rating scales for depression and to evaluate these constructs across clinical trials of a rapid-acting antidepressant (ketamine).
Methods: Exploratory factor analysis (EFA) was conducted on baseline ratings from the Beck Depression Inventory (BDI), the Hamilton Depression Rating Scale (HAM-D), the Montgomery-Asberg Depression Rating Scale (MADRS), and the Snaith-Hamilton Pleasure Rating Scale (SHAPS). Inpatients with major depressive disorder (n = 76) or bipolar depression (n = 43) were participating in clinical ketamine trials. The trajectories of the resulting unidimensional scores were evaluated in 41 subjects with bipolar depression who participated in clinical ketamine trials.
Results: The best solution, which exhibited excellent fit to the data, comprised eight factors: Depressed Mood, Tension, Negative Cognition, Impaired Sleep, Suicidal Thoughts, Reduced Appetite, Anhedonia, and Amotivation. Various response patterns were observed across the clinical trial data, both in treatment effect (ketamine versus placebo) and in degree of placebo response, suggesting that use of these unidimensional constructs may reveal patterns not observed with traditional scoring of individual instruments.
Limitations: Limitations include: 1) small sample (and related inability to confirm measurement invariance); 2) absence of an independent sample for confirmation of factor structure; and 3) the treatment-resistant nature of the population, which may limit generalizability.
Conclusions: The empirical identification of unidimensional constructs creates more refined scores that may elucidate the connection between specific symptoms and underlying pathophysiology.

1. Introduction

Under DSM-5 criteria, an estimated 227 combinations of symptoms will lead to a diagnosis of a depressive episode. As a result, a wide range of individuals who meet criteria for depression may overlap on only a limited number of symptoms (Ostergaard et al., 2011; Zimmerman et al., 2015). Indeed, the heterogeneity inherent in the diagnosis of major depressive disorder (MDD) has been a consistent obstacle for identifying viable depression-specific biomarkers that could signal the presence of the disorder as well as predict and track treatment response

(Leuchter et al., 2010; Zarate et al., 2013).

Isolating specific clusters of the depressive syndrome with a particular biological signature may be an important step towards advancing translational research into depression and, concomitantly, developing novel therapeutics. However, the depression rating scales commonly used in clinical trials survey a variety of symptoms that reflect DSM criteria, which limits research in several key ways. For instance, such rating scales are useful in dichotomizing individuals into depressed vs. non-depressed samples, but provide little insight into specific symptom clusters that would lead to more homogeneous subgroups, as advocated

by efforts such as the NIMH RDoC (Woody and Gibb, 2015). In this context, using unidimensional depressive symptom constructs could reduce variability in the data and increase the precision of attempts to connect specific symptoms with pathophysiology. However, it can be difficult to translate the multifaceted construct of depression across modalities—that is, from depressed patients to healthy control samples or to preclinical models. For example, a cross-method translational approach might first involve isolating a particular symptom construct (e.g, anhedonia or approach motivation) into specific neural circuits in patient samples, followed by an experimental paradigm to induce anhedonic symptoms in non-depressed healthy control participants, and finally into preclinical models of anhedonia in animal studies (Treadway and Zald, 2011). In a similarly translational fashion, findings from preclinical models of anhedonia could have implications for both healthy control and patient samples. However, this approach may be unnecessarily complicated by use of diffuse constructs like 'depression'. Moreover, depression symptom domains may not have uniform response to treatment. For example, some symptom clusters may be particularly vulnerable to the placebo effect, some may exhibit differential response latency, and others still may not respond to a given intervention. These properties may have unexpected effects on the efficiency and precision of clinical trials, and it is possible—even likely—that researchers are unnecessarily handicapped by redundant use of multidimensional outcome measures.

This analysis sought to identify the unidimensional constructs measured by commonly used rating scales of depression and anhedonia, including the Beck Depression Inventory (BDI), the Hamilton Depression Rating Scale (HAM-D), the Montgomery-Asberg Depression Rating Scale (MADRS), and the Snaith-Hamilton Pleasure Rating Scale (SHAPS). Identifying such unidimensional constructs could then inform the identification of neurobiologically distinct subtypes (also known as biotypes) of depression. In particular, the inclusion of anhedonia and cognitive symptom-specific measures of depression across both clinician-administered and self-report assessments would allow the comprehensive examination of a range of experiences associated with depression. As an initial demonstration of these unidimensional constructs, and in order to assess whether the identified constructs have neurobiological relevance, we examined how these symptoms change in response to a rapid-acting pharmacologic intervention (the glutamatergic modulator ketamine) compared with traditional measures of depression. The literature on depressive biotypes is growing rapidly—in part related to imaging connectivity analyses (Drysdale et al., 2017; Williams, 2017) and the ongoing search for central or peripheral biomarkers (Lamers et al., 2013)—and we believe that careful parcellation of depressive symptoms and behaviors is critical to ensuring that these biotypes have clinical relevance and significance.

2. Methods

2.1. Participants

One hundred nineteen currently depressed patients (61 male, 58 female; aged 21–66, mean age = 45.28 years, SD = 12.45) were recruited from inpatient studies conducted at the National Institute of Mental Health, National Institutes of Health (NIMH-NIH), Bethesda, MD, USA. The patient sample comprised 76 subjects diagnosed with major depressive disorder (MDD) and 43 diagnosed with bipolar depression (either I or II); the presence of psychotic features was an exclusion criterion for both diagnoses. All patients participated in trials on the same research unit and were assessed and treated by the same clinical and research staff. All trials examined the use of ketamine as a rapid-acting antidepressant; results have been previously published (Diazgranados et al., 2010b; Ibrahim et al., 2012; Zarate et al., 2012, 2006).

Participants were initially screened and evaluated for eligibility for research participation, which included an initial MADRS score ≥ 20 or

a HAM-D score ≥ 18 across all trials. Once at the NIH, diagnosis was established using the Structured Clinical Interview for Axis I Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV (First et al., 2002)) and corroborated by a team of clinicians using all available information. All subjects were in good physical health as determined by medical history, physical examination, and laboratory tests. Exclusion criteria included pregnancy, nursing, or illicit comorbid substance abuse in the previous three months. Written informed consent was obtained from all participants, and the NIH combined Neuroscience Institute Review Board approved the study.

Across all trials, the co-occurrence of Axis I anxiety disorders was permitted if it was not the primary focus of treatment within the past 12-month period. At hospital admission, all subjects were currently experiencing a major depressive episode lasting at least four weeks. Once admitted and where necessary to comply with individual protocols, subjects were tapered off of their existing medications and underwent a two-week drug-free period (five weeks for fluoxetine, three weeks for aripiprazole) before study baseline. All patients diagnosed with bipolar depression were maintained on a therapeutic dose of either lithium (serum lithium, 0.6–$1.2\,\mathrm{mEq\,l^{-1}}$) or valproic acid ($50$–$125\,\mathrm{\mu g\,ml^{-1}}$) for four weeks without exhibiting an antidepressant response to the prescribed medication. No other psychotropic medication or psychotherapy was permitted during the drug-free period prior to study baseline or throughout the study. All subjects, with one exception, were considered treatment-resistant, defined as having failed to respond at least one adequate treatment trial, as determined by the Antidepressant Treatment History Form (Sackeim, 2001).

2.2. Design

Details regarding study designs can be found elsewhere (Diazgranados et al., 2010b; Ibrahim et al., 2012; Zarate et al., 2012, 2006). Briefly, patients were administered psychiatric scales in the morning, approximately one hour before beginning their first infusion (regardless of whether the study was open-label or placebo-controlled). This pre-infusion baseline was a time where patients had been medication-free for at least two weeks, with the exception of those patients with bipolar depression who were maintained on lithium or valproate. Ketamine was administered intravenously at 0.5 mg/kg; in the placebo-controlled studies, saline infusions were used as the control condition. The psychiatric rating scales were re-administered to patients at 40, 80, 120, and 230 min post-infusion and at Days 1, 2, and 3.

From the larger patient group of 119 depressed participants, longitudinal data from 41 subjects with bipolar depression were used to assess the unidimensional scores in clinical trials. Most of the bipolar depression patients (n = 33) had participated in one of two randomized, placebo-controlled, crossover trials of ketamine (an initial trial and a replication) (Diazgranados et al., 2010b; Zarate et al., 2012). The remaining eight participants were drawn from ongoing biomarker studies. These studies were specifically selected for use in this preliminary analysis due to the uniformity of diagnosis, use of all relevant measures, and similarity of research design; it should be noted that for the eight participants drawn from our ongoing biomarker studies, identical methods were used regarding recruitment procedures, inclusion/exclusion criteria, and study protocols. Patient demographics and treatment response did not differ across sources (see Supplementary Table 1).

2.3. Measures

The BDI (Beck et al., 1961) is a 21-item self-reported measure of depression severity. Items are framed as aspects of depressive symptomology such as "Sadness". Answers are measured on a 0–3 scale, with higher scores indicating increased severity of depressive symptoms (e.g., "I do not feel sad" to "I am so sad or unhappy I can't stand it"). The BDI has high internal reliability (Beck et al., 1961) and concurrent

validity with the HAM-D (Beck et al., 1974).

The HAM-D (Hamilton, 1960) is a 17-item clinician-administered scale that assesses severity of depression. As with the BDI, a range of depressive symptomology is assessed; answers range from 0 to 4, and higher scores indicate increased severity of depressive symptoms (e.g., for the item "Insomnia: Early in the Night", answers range from "No difficulty" to "Complains of nightly difficulty falling asleep").

The MADRS (Montgomery and Asberg, 1979) is a 10-item clinician-administered scale. Ratings of depressive symptomology are made on a scale of 0–6, with higher scores indicating increased severity of depressive symptoms (e.g., for the item "Reduced Appetite", answers range from "Normal or increased appetite" to "Needs persuasion to eat at all").

The SHAPS (Snaith et al., 1995) is a 14-item self-reported measure of anhedonia. Items assess anhedonia on a 1–4 scale ranging from "strongly agree" to "strongly disagree". An example of an item from the SHAPS is "I would find pleasure in my hobbies and pastimes". Higher scores indicate increased levels of anhedonia.

2.4. Statistical analysis

An exploratory factor analysis (EFA) using Mplus version 7.4 (Muthen and Muthen, 2012) was used to explore baseline data from the 119 participants. An EFA was selected rather than a principal components analysis because the former is the most appropriate method for identifying unidimensional constructs, whereas the latter is most appropriate for data reduction.

First, data were prepared by collapsing response categories with fewer than 10 respondents (e.g., if a BDI item score of 2 was endorsed by only seven participants, those participants were pooled with those who received a score of 1). Second, items with insufficient variability were deleted (e.g., items where fewer than 10 individuals received a non-zero score, which eliminated two items from the HAM-D). Third, given the substantial overlap in some items from different scales, the correlation matrix of all items was examined for values greater than 0.90, and one item was removed (e.g., the MADRS reduced appetite item was retained, but the HAM-D reduced appetite item was excluded; this occurred for two additional HAM-D items). Finally, an EFA with promax (oblique) rotation was performed on the polychoric correlation matrix, with solutions of 1 through 10 factors. Model fit was evaluated using standard interpretation (Hu and Bentler, 1999) of common fit statistics: the root mean square error of approximation (RMSEA; values < 0.05 considered good), the comparative fit index (CFI; values > 0.95 considered good), and the Tucker-Lewis fit index (TLI, values > 0.95 considered good). The chi-square statistic for improvement in model fit was also evaluated.

Using the best model from the EFA stage, unidimensional constructs (i.e., equally weighted item totals) were created. First, items were selected onto factors where the loading was statistically significant and greater than 0.30; if an item loaded significantly onto more than one factor, the factor with the strongest loading was selected. Second, items from the different measures were put onto the same scale by dividing the score by the "points possible" on the item (i.e., a score of 1 on a 0-to-3 scale was transformed to 0.33). Finally, to facilitate comparison across unidimensional scores, the scaled item scores were summed and divided by the number of items to generate an item-mean score for each unidimensional construct.

The item-mean unidimensional scores over the course of the crossover trial were analysed using a linear mixed model with restricted maximum likelihood estimation and a compound symmetry covariance structure (selected based on fit indices, calculated separately for drug condition), a random subject effect, and fixed effects of time, drug, and their interaction. The period-specific baseline value was entered as a covariate. Effect sizes with 95% confidence intervals were calculated using the least-square mean estimates of difference (and 95% CI) and associated degrees of freedom between ketamine and placebo

Table 1
Participant demographics.

	All participants n = 119 (MDD and BD)		Longitudinal data n = 41 (Bipolar Depression only)	
	Mean	**SD**	**Mean**	**SD**
Age	45.28	12.45	46.38	10.73
Age of onset of first depressive episode	18.82	10.05	18.03	7.10
Length of current episode (months)	54.43	94.44	17.87	20.68
BMI	29.56	6.53	29.96	5.81
Baseline BDI	27.97	7.97	29.32	7.65
Baseline HAM-D	21.40	4.18	21.98	4.22
Baseline MADRS	33.51	4.80	33.90	5.14
Baseline SHAPS	37.95	6.72	36.90	7.75
	n	*%*	*n*	*%*
Female	58	49	23	44
Caucasian	106	89	34	85
Bipolar Depression	43	36	41	100
History of suicide attempt	47	40	18	46

Abbreviations: BDI: Beck Depression Inventory; HAM-D: Hamilton Rating Scale for Depression; MADRS: Montgomery-Åsberg Depression Rating Scale; SHAPS: Snaith-Hamilton Pleasure Scale; BMI: Body Mass Index, MDD: major depressive disorder.
Note: All patients whose longitudinal data were used had bipolar depression and were maintained on therapeutic doses of lithium or valproic acid during the study. These patients were selected for a sample longitudinal analysis due to the consistency of scale administration and the similarity of their study designs. One person was missing BMI, and two people were missing data on history of suicide attempt. Longitudinal data were drawn from three identically designed studies (n = 18, n = 15, and n = 8). Participant demographics did not differ by study source (see Supplementary Table 1).

conditions. These analyses combined multiple trials and were purely exploratory. For that reason, no adjustments for multiple comparisons were made.

3. Results

Patient demographics and baseline scores on psychiatric scales are reported in Table 1. Patients were moderately to severely depressed, as indicated by an average MADRS score of 34 and an average BDI score of 28.

Results of the EFA indicated that an eight-factor solution was the best fit to the data (Table 2). The fit of the model to the data was good; the RMSEA was 0.025 (95% CI: 0.008–0.036), the CFI was 0.97, and the TLI was 0.95. Factor correlations were negligible to moderate and non-significant in most cases. We labelled the subsequent unidimensional constructs as: *Depressed Mood, Tension, Negative Cognition, Impaired Sleep, Suicidal Thoughts, Reduced Appetite, Anhedonia,* and *Amotivation*; all were of variable size (range three to 12 items, see Table 3). The summed item-mean scores, which were used in subsequent analyses, were strongly related to the factor scores (Pearson correlations ranging from $r = 0.84$ to $r = 0.98$, all $p < 0.0001$) (see Supplementary Table S3 and Supplementary Fig. S1).

Using the longitudinal data (n = 41), we then evaluated the effect of ketamine on both the unidimensional scores and the conventional instrument scores originally used in the trials (BDI, MADRS, HAM-D, and SHAPS) (Table 4). Trajectories of change are shown in Fig. 1. With the exceptions of *Tension, Impaired Sleep,* and *Reduced Appetite*, a significant effect ($p < 0.05$) of ketamine relative to placebo at all post-baseline timepoints was observed on both the unidimensional scores and the conventional rating instrument scores. The effect of ketamine on *Tension* was significant at all timepoints except Day 3. The effect of ketamine on *Impaired Sleep* was not significant until 230 min. The effect of drug on *Reduced Appetite* reversed midway through the study, such that the ketamine condition was significantly worse at timepoints 40, 80,

Table 2
Results of exploratory factor analysis (n = 119).

Number of factors	Model parameters	X^2	RMSEA (0.06)	RMSEA 90% CI	CFI (0.95)	TLI (0.95)	SRMR (0.08)	X^2 (p) for fit improvement (< 0.05)
1	57	2420.38	0.07	(0.064–0.075)	0.65	0.64	0.15	n/a
2	113	1923.23	0.05	(0.043–0.056)	0.83	0.81	0.12	< 0.0001
3	168	1765.78	0.05	(0.037–0.051)	0.87	0.85	0.10	< 0.0001
4	222	1618.75	0.04	(0.030–0.046)	0.90	0.89	0.09	< 0.0001
5	275	1511.97	0.04	(0.025–0.043)	0.92	0.91	0.09	< 0.0001
6	327	1413.98	0.03	(0.019–0.040)	0.94	0.93	0.08	< 0.0001
7	378	1338.81	0.03	(0.016–0.038)	0.95	0.94	0.07	0.0035
8	428	1257.14	0.03	(0.008–0.036)	0.97	0.95	0.07	0.0003
9	477	1187.82	0.02	(0–0.034)	0.97	0.96	0.06	0.0175
10	525	1133.41	0.02	(0–0.034)	0.98	0.96	0.06	0.1069

Note: RMSEA: root mean square error of approximation; CFI: comparative fit index; TLI: Tucker-Lewis fit index; SRMR: standardized root mean square residual. The cut-off for each fit statistic, or the value to which a given fit statistic should be close to in order to signal "good" fit (Hu and Bentler, 1999), is listed in each column header; these are rules of thumb. Chi-square is scaled to reflect the use of the mean- and variance-adjusted weighted least squares estimator.

and 120 min, but not different from placebo at 230 min, Day 1, and Day 2. Taken together, these results suggest that unique symptom clusters associated with treatment response to ketamine may be more readily identified using statistically identified unidimensional constructs than conventional rating instruments.

Although the significant effect of drug was apparent across both unidimensional and conventional scores, the magnitude of the effect varied widely. Using Day 1 for the purposes of illustration (Fig. 2), we noted that effect sizes for ketamine versus placebo ranged from very small and non-significant (*Reduced Appetite*: $d = 0.20$, 95% CI: − 0.084 to 0.48) to large (*Amotivation*: $d = 1.45$, 95% CI: 0.94–1.96). Less variability was observed in the conventional rating instrument scores, which all exceeded 1.0.

Degree of improvement on placebo (i.e., change from baseline during the placebo condition) also varied. For the purposes of modelling placebo response, baseline scores were moved from covariate to dependent variable so later timepoints could be directly compared to baseline, and the models were re-run. Three of the conventional rating instrument scores exhibited significant placebo response (i.e., change from baseline) at 40 min: BDI ($d = 0.76$, 95% CI: 0.49–1.03), HAM-D ($d = 0.57$, 95% CI: 0.30–0.84), and MADRS ($d = 0.68$, 95% CI: 0.40–0.95). In contrast, five of the eight unidimensional scores exhibited significant placebo response at 40 min: *Depressed Mood* ($d = 0.69$, 95%CI 0.42–0.96), *Tension* ($d = 0.55$, 95% CI: 0.28–0.82), *Negative Cognition* ($d = 0.64$, 95% CI: 0.37–0.92), *Suicidal Thoughts* ($d = 0.43$, 95% CI: 0.16–0.70), and *Amotivation* ($d = 0.60$, 95% CI: 0.32–0.87). With the exceptions of *Amotivation* and *Suicidal Thoughts*, these placebo effects remained significant (though attenuated) at Day 1.

Table 3
Eight-factor solution (n = 119).

Source	Item	Loading	Source	Item	Loading
Depressed mood			**Impaired sleep**		
MADRS	Reported Sadness	0.79	MADRS	Reduced Sleep	0.72
HAM-D	Depressed Mood	0.84	HAM-D	Early Insomnia	0.70
MADRS	Concentration Difficulties	0.37	HAM-D	Middle Insomnia	0.62
MADRS	Inability to Feel	0.61	HAM-D	Late Insomnia	0.58
HAM-D	Inability to Work	0.54	BDI	Insomnia	0.65
MADRS	Apparent Sadness	0.77	**Reduced appetite**		
HAM-D	Psychomotor Retardation	0.55	MADRS	Reduced Appetite	0.92
BDI	Sadness	0.51	HAM-D	Weight Loss	0.49
Tension			BDI	Reduced Appetite	0.76
HAM-D	Psychological Anxiety	0.90	SHAPS	Reduced Enjoyment Eating	0.53
MADRS	Inner Tension	0.86	**Anhedonia**		
HAM-D	Hypochondriasis	0.32	SHAPS	Reduced Enjoyment TV/Radio	0.60
BDI	Fatigue	0.30	SHAPS	Reduced Enjoyment w/Family	0.43
Negative cognition			SHAPS	Reduced Enjoyment Hobbies	0.41
MADRS	Pessimistic Thoughts	0.39	SHAPS	Reduced Enjoyment Smells	0.60
HAM-D	Guilt	0.61	SHAPS	Reduced Enjoyment Others' Happiness	0.72
BDI	Thoughts of Failure	0.49	SHAPS	Reduced Enjoyment Appearance	0.80
BDI	Guilt	0.71	SHAPS	Reduced Enjoyment Reading	0.58
BDI	Feelings of Punishment	0.48	SHAPS	Reduced Enjoyment Tea/Coffee	0.66
BDI	Disappointment in Self	0.53	SHAPS	Reduced Enjoyment Small Pleasures	0.75
BDI	Self-Criticism	0.74	SHAPS	Reduced Enjoyment Landscape	0.62
BDI	Increased Crying	0.59	SHAPS	Reduced Enjoyment Helping Others	0.57
BDI	Worthlessness	0.70	SHAPS	Reduced Enjoyment Receiving Praise	0.86
BDI	Reduced Sexual Interest		**Amotivation**		
Suicidal thoughts			BDI	Dissatisfaction with Life	0.57
MADRS	Suicidal Thoughts	0.70	BDI	Loss of Interest in People	0.33
BDI	Hopelessness	0.67	BDI	Indecisiveness	0.34
BDI	Suicidal Thoughts	0.78	BDI	Inability to Work	0.72
			BDI	Irritability	0.62

BDI: Beck Depression Inventory; HAM-D: Hamilton Rating Scale for Depression; MADRS: Montgomery-Åsberg Depression Rating Scale; SHAPS: Snaith-Hamilton Pleasure Scale; BMI: Body Mass Index.
Note: Items were selected onto subscales when the standardized loading was statistically significant ($p < 0.05$) and greater than 0.30; if an item loaded significantly onto more than one factor, the factor with the strongest loading was selected. Eight items (three from the BDI, three from the HAM-D, one from the MADRS, and one from SHAPS) were excluded, leaving 51 items across the eight subscales. The full loading table is shown in Supplementary Table 2.

Table 4
Type III tests of fixed effects (n = 41).

	Drug F (p)[a]	Minutes F (p)[b]	Drug * Minutes F (p)[b]
Constructs			
Depressed Mood	43.40 (< 0.0001)	3.30 (0.004)	2.00 (0.07)
Tension	41.22 (< 0.0001)	9.02 (< 0.0001)	3.31 (0.004)
Negative Cognition	48.90 (< 0.0001)	9.8 (< 0.0001)	3.08 (0.006)
Impaired Sleep	13.19 (0.0009)	1.45 (0.194)	0.30 (0.94)
Suicidal Thoughts	47.96 (< 0.0001)	9.42 (< 0.0001)	4.32 (0.0003)
Reduced Appetite	0.59 (0.44)	2.78 (0.01)	7.33 (< 0.0001)
Anhedonia	18.07 (0.0001)	1.99 (0.07)	0.71 (0.64)
Amotivation	38.50 (< 0.0001)	3.83 (0.001)	0.55 (0.77)
Conventional scores			
BDI Total	44.19 (< 0.0001)	8.10 (< 0.0001)	1.38 (0.22)
HAM-D Total	36.09 (< 0.0001)	4.68 (0.0001)	1.22 (0.30)
MADRS Total	47.37 (< 0.0001)	3.90 (0.0009)	1.62 (0.14)
SHAPS Total	15.41 (0.0003)	2.32 (0.03)	0.77 (0.60)

Note: BDI: Beck Depression Inventory; HAM-D: Hamilton Rating Scale for Depression; MADRS: Montgomery-Asberg Depression Rating Scale; SHAPS: Snaith Hamilton Pleasure Scale; DF: degrees of freedom. Results of a mixed model with compound symmetry covariance structure for repeated measures, random subject effect, and Satterthwaite's approximation for degrees of freedom. Period-specific baseline was entered as a covariate.

[a] Numerator DF was 1, denominator DF ranged from 31.3 to 57.2.

[b] Numerator DF was 6, denominator DF ranged from 290 to 384.

4. Discussion

This EFA of the MADRS, HAM-D, BDI, and SHAPS was conducted with ratings obtained from subjects with treatment–resistant MDD and bipolar depression in an effort to identify and reduce heterogeneity within depression rating scales. The scales were found to measure several unidimensional constructs, which we identified as: *Depressed Mood, Tension, Negative Cognition, Impaired Sleep, Suicidal Thoughts, Reduced Appetite, Anhedonia,* and *Amotivation.* The *Anhedonia* construct comprised items from the SHAPS, whereas the *Negative Cognition*
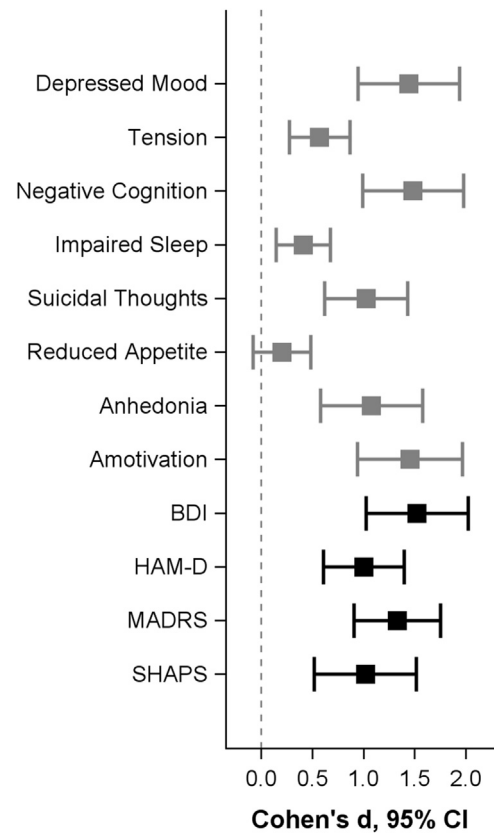
Fig. 2. Effect sizes (ketamine versus placebo) at Day 1 in the Bipolar Depression analysis (n = 41). Effect sizes were calculated using least square mean difference estimates, and associated degrees of freedom, at Day 1 from the mixed model. BDI: Beck Depression Inventory; HAM-D: Hamilton Rating Scale for Depression; MADRS: Montgomery-Asberg Depression Rating Scale; SHAPS: Snaith Hamilton Pleasure Scale.
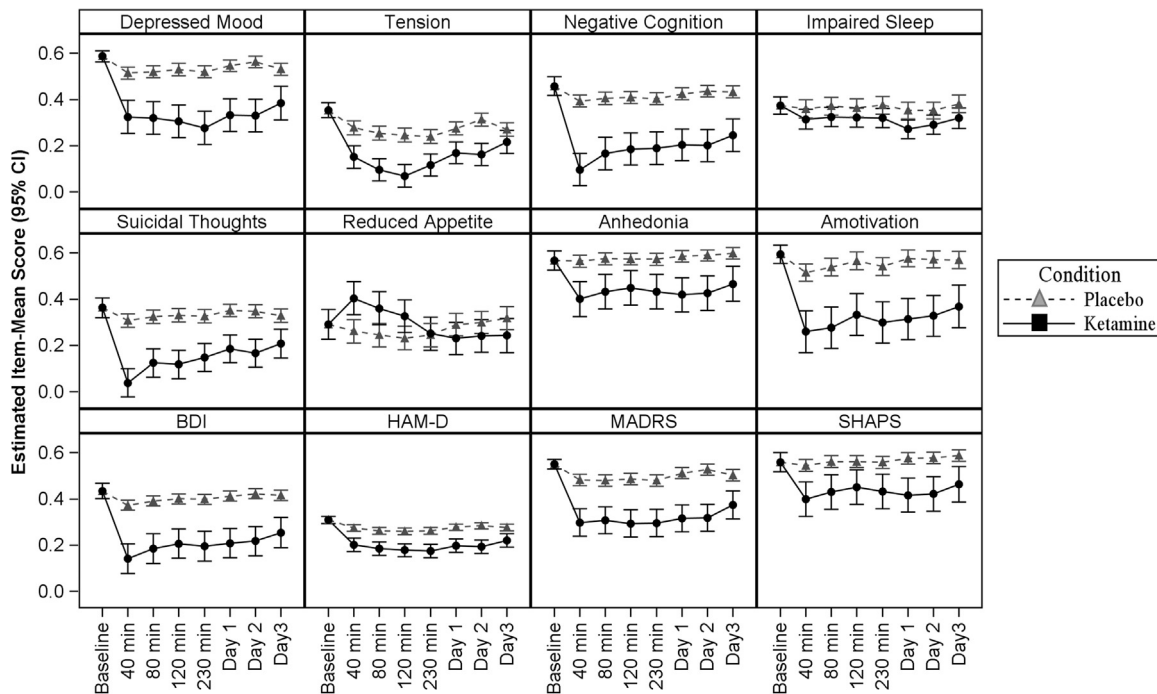
Fig. 1. Results of mixed models in Bipolar Depression analysis (n = 41). Item-mean scores reflect the average proportion of points endorsed to points available across items on the subscale. For all constructs except *Tension, Impaired Sleep,* and *Reduced Appetite,* the difference between ketamine and placebo was statistically significant ($p < 0.05$) at all post-baseline assessments. The effect of drug on *Tension* was significant except at Day 3. The effect of drug on *Impaired Sleep* was not significant until 230 min. The effect of drug on *Reduced Appetite* was significant (in favour of placebo) until 230, after which it was not significant. BDI: Beck Depression Inventory; HAM-D: Hamilton Rating Scale for Depression; MADRS: Montgomery-Asberg Depression Rating Scale; SHAPS: Snaith Hamilton Pleasure Scale.

construct was primarily composed of items from the BDI, highlighting the importance of using a range of clinical and self-report measures of both depressive and anhedonic symptoms to obtain a more comprehensive picture of depressive symptomatology.

As a first demonstration of the use of these constructs, we re-analyzed existing clinical trial data using the unidimensional scores as outcomes. The effect of ketamine varied from small to moderate across the unidimensional scores, demonstrating that the depression symptom domains did not respond uniformly to ketamine administration. In contrast, the effect sizes across total scores from the conventional measures (MADRS, HAM-D, BDI, and SHAPS) were uniformly moderate. Furthermore, variable patterns of placebo response were observed; five of the eight unidimensional scores were significantly improved at 40 min from baseline, while three of the four conventional measures were improved after placebo. Overall, the results of the EFA indicated that several unidimensional constructs can be measured by widely used depression rating scales. In addition, our preliminary evaluation of clinical trial data in inpatients with treatment-resistant MDD and bipolar depression suggests that using these unidimensional scores may reveal response trajectories that were previously obscured by use of heterogeneous conventional rating instruments.

Our study extends the work of other researchers who identified homogenous constructs that could be measured by depression and anhedonia rating scales by widening the EFA to include constructs across—rather than within—scales. In one respect, the resulting EFA is quite similar to other factor analyses of MDD symptoms (Vrieze et al., 2014), including analyses of the HAM-D6 as a distilled measure of melancholia (Bech et al., 2011; Ostergaard et al., 2014), and other evaluations of the SHAPS as a unidimensional scale with divergent validity from other depression rating measures (Nakonezny et al., 2010, 2015). Similarly, Grunebaum and colleagues, in their separate factor analyses of the HAM-D and BDI, identified constructs that reflect psychic depression, loss of motivated behavior, disturbed thinking, anxiety, and sleep disturbance from the HAM-D, and subjective depression, self-blame, and somatic complaints from the BDI (Grunebaum et al., 2005). These subscale scores have also been linked to distinct neuroimaging findings using PET, as well as response to treatment (Grunebaum et al., 2013; Milak et al., 2005). The present study builds upon this work by combining scales that represent all facets of depressive symptoms, thus enabling us to more holistically model their heterogeneity. Although we excluded redundant items, many items remained that were similar across scales. This allowed us to identify factors represented by few items (for instance, suicidal ideation) that, traditionally, would have been assessed by only one item in each scale, resulting in greater measurement bias.

The variable effects of ketamine observed across unidimensional scores are consistent with previously published findings of ketamine's rapid and broad antidepressant, anti-anxiolytic, anti-anhedonic, and anti-suicidal effects (Diazgranados et al., 2010a; Feder et al., 2014; Lally et al., 2014). Previous analyses of these data revealed that ketamine had observable effects on both anhedonia and suicidal ideation as well as on neural correlates identified via PET imaging, and that these effects occurred independently from its overall effect on depressive symptoms (Ballard et al., 2014a, 2014b; Lally et al., 2014, 2015). It is likely that the PET imaging findings were most closely related to the constructs assessed by the *Depressed Mood*, *Anhedonia*, and *Suicidal Ideation* factors. Similarly, a recent meta-analysis on functional connectivity suggested that depressed and anxious patients exhibited a number of depressive and anxious biotypes (Williams, 2017). We suspect that these biotypes may be more easily validated with unidimensional constructs like those proposed here than with conventional, heterogeneous measures. For example, the rumination biotype (Williams, 2017), characterized by default mode hyper-activation, could be linked to the *Negative Cognition* construct. In turn, the apprehension biotype, which is characterized by hypoactivation in the salience circuit, could be linked to the *Tension* construct, and the

anhedonia biotype, characterized by reward circuit hypoactivation, could be compared to both the *Anhedonia* and *Amotivation* constructs. In addition, Drysdale et al. (2017) developed a series of fMRI-based biotypes derived from frontostriatal and limbic connectivity and found that these differed on measures of anxiety, anhedonia, and psychomotor agitation, providing additional evidence that the constructs identified in the present study might fit into a dimensional approach that could be used to assess the neurobiology of depression, as advocated by efforts such as the NIMH RDoC (Cuthbert, 2014).

A major strength of this approach is that data obtained from multiple trials conducted by the same clinical researchers on the same clinical inpatient unit were combined, which reduces variability in the administration and interpretation of the original rating instruments. However, the study is also associated with several limitations. These include: 1) the small sample size; 2) the absence of an independent sample for confirmation of the EFA; 3) the fact that the participant sample comprised individuals with treatment-resistant MDD or bipolar depression, which might not translate across the continuum of individuals with mood disorders or healthy individuals; 4) the possibility that instruments such as the Temporal Experience of Pleasure Scale, which makes a clear distinction between anticipatory and consummatory anhedonia (Gard et al., 2006), could ultimately be more useful in future analyses than the SHAPS; and 5) that while these factors somewhat align with biotypes as proposed in the neuroimaging literature, future analyses should explicitly link these factors to functional connectivity data, experimental paradigms, plasma markers, and treatment outcomes. Finally, the small sample size precluded our ability to assess measurement invariance across diagnosis (MDD versus bipolar disorder) or over time. The possibility that the measurement properties of the factor model may differ across disorders or across timepoints is a limitation that must be addressed in future analyses.

5. Conclusions

As interest in identifying biotypes of subjects with mood disorders develops, so will the need for appropriate and specific clinical assessments. The present study has underscored that factor analysis techniques can identify unidimensional constructs from a range of depressive symptoms. By using these unidimensional scores instead of total measure scores, we may be able to increase the precision with which we clinically describe these new and emerging biotypes, as well as measure response to treatment. Such an approach may be an important step towards parsing the heterogeneity of depressive symptoms.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2018.01.027.

References

Ballard, E.D., Ionescu, D.F., Vande Voort, J.L., Niciu, M.J., Richards, E.M., Luckenbaugh,

D.A., Brutsche, N.E., Ameli, R., Furey, M.L., Zarate Jr., C.A., 2014a. Improvement in suicidal ideation after ketamine infusion: relationship to reductions in depression and anxiety. J. Psychiatr. Res. 58, 161–166.

Ballard, E.D., Lally, N., Nugent, A.C., Furey, M.L., Luckenbaugh, D.A., Zarate Jr., C.A., 2014b. Neural correlates of suicidal ideation and its reduction in depression. Int. J. Neuropsychopharmacol. 18.

Bech, P., Fava, M., Trivedi, M.H., Wisniewski, S.R., Rush, A.J., 2011. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. J. Affect. Disord. 132, 396–400.

Beck, A.T., Rial, W.Y., Rickets, K., 1974. Short form of depression inventory: cross-validation. Psychol. Rep. 34, 1184–1186.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561–571.

Cuthbert, B.N., 2014. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 13, 28–35.

Diazgranados, N., Ibrahim, L., Brutsche, N.E., Ameli, R., Henter, I.D., Luckenbaugh, D.A., Machado-Vieira, R., Zarate, C.A., 2010a. Rapid resolution of suicidal ideation after a single infusion of an N-methyl-D-aspartate antagonist in patients with treatment resistant major depressive disorder. J. Clin. Psychiatry 71, 1605–1611.

Diazgranados, N., Ibrahim, L., Brutsche, N.E., Newberg, A., Kronstein, P., Khalife, S., Kammerer, W.A., Quezado, Z., Luckenbaugh, D.A., Salvadore, G., Machado-Vieira, R., Manji, H.K., Zarate Jr., C.A., 2010b. A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. Arch. Gen. Psychiatry 67, 793–802.

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., Schatzberg, A.F., Sudheimer, K., Keller, J., Mayberg, H.S., Gunning, F.M., Alexopoulos, G.S., Fox, M.D., Pascual-Leone, A., Voss, H.U., Casey, B.J., Dubin, M.J., Liston, C., 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23, 28–38.

Feder, A., Parides, M.K., Murrough, J.W., Perez, A.M., Morgan, J.E., Saxena, S., Kirkwood, K., aan het Rot, M., Lapidus, K.A.B., Wan, L.-B., Iosifescu, D., Charney, D.S., 2014. Efficacy of intravenous ketamine for treatment of chronic posttraumatic stress disorder. A randomized clinical trial. JAMA Psychiatry 71, 681–688.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P). Biometrics Research. New York State Psychiatric Institute, New York.

Gard, D.E., Gard, M.G., Kring, A.M., John, O.P., 2006. Anticipatory and consummatory components of the experience of pleasure: a scale development study. J. Res. Personal. 40, 1086–1102.

Grunebaum, M.F., Keilp, J., Li, S., Ellis, S.P., Burke, A.K., Oquendo, M.A., Mann, J.J., 2005. Symptom components of standard depression scales and past suicidal behavior. J. Affect. Disord. 87, 73–82.

Grunebaum, M.F., Keilp, J.G., Ellis, S.P., Sudol, K., Bauer, N., Burke, A.K., Oquendo, M.A., Mann, J.J., 2013. SSRI versus bupropion effects on symptom clusters in suicidal depression: post hoc analysis of a randomized clinical trial. J. Clin. Psychiatry 74, 872–879.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

Hu, Lt, Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55.

Ibrahim, L., Diazgranados, N., Franco-Chaves, J., Brutsche, N., Henter, I.D., Kronstein, P., Moaddel, R., Wainer, I., Luckenbaugh, D.A., Manji, H.K., Zarate Jr., C.A., 2012. Course of improvement in depressive symptoms to a single intravenous infusion of ketamine vs add-on riluzole: results from a 4-week, double-blind, placebo-controlled study. Neuropsychopharmacology 37, 1526–1533.

Lally, N., Nugent, A.C., Luckenbaugh, D.A., Ameli, R., Roiser, J.P., Zarate, C.A., 2014. Anti-anhedonic effect of ketamine and its neural correlates in treatment-resistant bipolar depression. Transl. Psychiatry 4, e469.

Lally, N., Nugent, A.C., Luckenbaugh, D.A., Niciu, M.J., Roiser, J.P., Zarate Jr., C.A., 2015. Neural correlates of change in major depressive disorder anhedonia following open-label ketamine. J. Psychopharmacol. 29, 596–607.

Lamers, F., Vogelzangs, N., Merikangas, K.R., de Jonge, P., Beekman, A.T., Penninx, B.W., 2013. Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol. Psychiatry 18, 692–699.

Leuchter, A.F., Cook, I.A., Hamilton, S.P., Narr, K.L., Toga, A., Hunter, A.M., Faull, K., Whitelegge, J., Andrews, A.M., Loo, J., Way, B., Nelson, S.F., Horvath, S., Lebowitz, B.D., 2010. Biomarkers to predict antidepressant response. Curr. Psychiatry Rep. 12, 553–562.

Milak, M.S., Parsey, R.V., Keilp, J., Oquendo, M.A., Malone, K.M., Mann, J.J., 2005. Neuroanatomic correlates of psychopathologic components of major depressive disorder. Arch. Gen. Psychiatry 62, 397–408.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

Muthen, L.K., Muthen, B.O., 2012. Mplus User's Guide, Seventh ed. Muthen & Muthen, Los Angeles, CA.

Nakonezny, P.A., Carmody, T.J., Morris, D.W., Kurian, B.T., Trivedi, M.H., 2010. Psychometric evaluation of the Snaith-Hamilton pleasure scale in adult outpatients with major depressive disorder. Int. Clin. Psychopharmacol. 25, 328–333.

Nakonezny, P.A., Morris, D.W., Greer, T.L., Byerly, M.J., Carmody, T.J., Grannemann, B.D., Bernstein, I.H., Trivedi, M.H., 2015. Evaluation of anhedonia with the Snaith-Hamilton Pleasure Scale (SHAPS) in adult outpatients with major depressive disorder. J. Psychiatr. Res. 65, 124–130.

Ostergaard, S.D., Bech, P., Trivedi, M.H., Wisniewski, S.R., Rush, A.J., Fava, M., 2014. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the research domain criteria (RDoC). J. Affect. Disord. 163, 18–24.

Ostergaard, S.D., Jensen, S.O., Bech, P., 2011. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiatr. Scand. 124, 495–496.

Sackeim, H.A., 2001. The definition and meaning of treatment-resistant depression. J. Clin. Psychiatry 62 (Suppl. 16), S10–S17.

Snaith, R.P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., Trigwell, P., 1995. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br. J. Psychiatry 167, 99–103.

Treadway, M.T., Zald, D.H., 2011. Reconsidering anhedonia in depression: lessons from translational neuroscience. Neurosci. Biobehav. Rev. 35, 537–555.

Vrieze, E., Demyttenaere, K., Bruffaerts, R., Hermans, D., Pizzagalli, D.A., Sienaert, P., Hompes, T., de Boer, P., Schmidt, M., Claes, S., 2014. Dimensions in major depressive disorder and their relevance for treatment outcome. J. Affect. Disord. 155, 35–41.

Williams, L.M., 2017. Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: a theoretical review of the evidence and future directions for clinical translation. Depress. Anxiety 34, 9–24.

Woody, M.L., Gibb, B.E., 2015. Integrating NIMH research domain criteria (RDoC) into depression research. Curr. Opin. Psychol. 4, 6–12.

Zarate Jr., C.A., Brutsche, N.E., Ibrahim, L., Franco-Chaves, J., Diazgranados, N., Cravchik, A., Selter, J., Marquardt, C.A., Liberty, V., Luckenbaugh, D.A., 2012. Replication of ketamine's antidepressant efficacy in bipolar depression: a randomized controlled add-on trial. Biol. Psychiatry 71, 939–946 (Epub 2012 Jan 2031).

Zarate Jr., C.A., Mathews, D.C., Furey, M.L., 2013. Human biomarkers of rapid antidepressant effects. Biol. Psychiatry 73, 1142–1155.

Zarate Jr., C.A., Singh, J.B., Carlson, P.J., Brutsche, N.E., Ameli, R., Luckenbaugh, D.A., Charney, D.S., Manji, H.K., 2006. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. Arch. Gen. Psychiatry 63, 856–864.

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., Dalrymple, K., 2015. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr. Psychiatry 56, 29–34.