Journal of Psychiatric Research 116 (2019) 178–184

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Recall accuracy for the symptoms of a major depressive episode among clinical trial participants

Boadie W. Dunlop[a,*], Maria Granros[b], Amber Lechner[a], Tanja Mletzko-Crowe[a], Charles B. Nemeroff[c], Helen S. Mayberg[d,1], W. Edward Craighead[a,e,1]

[a] Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA
[b] Department of Psychology, University of Georgia, Athens, GA, USA
[c] Institute for Early Life Adversity Research, University of Texas Dell Medical School in Austin, Austin, TX, USA
[d] Departments of Neurology and Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[e] Department of Psychology, Emory University, Atlanta, GA, USA

ARTICLE INFO

Keywords:
Dysphoria
Anhedonia
Suicide
Depression
CBT
Antidepressant

ABSTRACT

For patients with major depressive disorder (MDD), approaches to treatment differ for those with a single versus recurrent episodes. Based on studies of community samples, however, accuracy is low for identifying past episodes. Recall accuracy among clinical samples with a well-defined major depressive episode (MDE) has not been examined previously. We evaluated episode recall accuracy in 79 MDD patients in follow-up of the Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study at 12- and 24-month timepoints after starting treatment. Using the Structured Clinical Interview for DSM-IV, patients were asked to recall whether they had been experiencing the nine criterion symptoms of an MDE at the time of their intake assessment. Accuracy of recall for the index MDE was high, with 95% of patients at month 12 and 85% at month 24 recalling sufficient symptoms to meet the diagnostic criteria. Recall accuracy for specific symptoms varied considerably, from > 90% for dysphoria and anhedonia, to 55% for psychomotor and weight/appetite changes. For the thoughts of death/suicide criterion, patients with erroneous recall were significantly more likely to recall having had the symptom at the intake evaluation (though they had denied it at the time) than vice versa ($p < .007$). Patients who have participated in a clinical trial are likely to recall accurately a past MDE up to two years prior. Optimal vigilance for suicidal ideation for treatment-naïve patients should include a combination of self-report and clinician assessments.

1. Introduction

Major depressive disorder (MDD) affects over 300 million people globally and is the largest contributor to global disability (World Health Organization, 2017). MDD can be diagnosed after a single major depressive episode (MDE), though most patients experience recurrent MDEs (Burcusa and Iacono, 2007). The odds of eventual MDE recurrence increase by 16% with each subsequent full episode (Solomon et al., 2000). Recurrence increases the negative consequences of depression, including suicide risk, workplace impairment, and economic and social burdens (Rohde et al., 2013). Recurrent MDD also affects treatment planning by increasing the need for maintenance psychotherapy or pharmacologic treatments to prevent future episodes (APA, 2010). Unfortunately, patients often struggle to accurately recall prior MDEs, yielding uncertainty around whether newly presenting patients are experiencing their first episode or have recurrent MDD.

Identification of recurrent MDD relies upon patients' remembering their prior MDEs. The accuracy of adults in recalling past MDEs has been examined in 9 studies, producing inconsistent results, with accuracy rates ranging from 33 to 73% (Table 1). These studies can be broadly assorted into those that inquire about a lifetime history of depression (LTH) versus those that ask about a specific depressive episode, referred to as the "index episode." Most prior studies used community samples; those that did not evaluated the recall accuracy of depressed patients who had sought treatment at some earlier time point. LTH has been criticized as an unstable categorical variable

* Corresponding author. Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 12 Executive Park Drive NE, 3rd Floor, Atlanta, GA, 30329, USA.
E-mail address: bdunlop@emory.edu (B.W. Dunlop).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.jpsychires.2019.03.008
Received 5 January 2019; Received in revised form 1 March 2019; Accepted 7 March 2019

Table 1
Studies assessing accuracy of recall of prior major depressive episodes.

Author	N	Time Between Assessments	Determination of Index Episode	Sample	Measures	Comparison	Percent Accurately Recalling Episode
Aneshensel et al. (1987)	214	4 years	Retrospective	Community	Single question probe	LTH MDD at Time 1–5 vs at Time 5	43
Warshaw et al. (1991)	536	6 years	Retrospective	Relatives of patients	SADS-L, RDC	LTH MDD Time 1 vs Time 2	73
Foley et al. (1998),[a]	562	5 years	Retrospective	Twins	SCID (adapted)	LTH MDD reported at index vs Time 2	51
Andrews et al. (1999)	27	25 years	Retrospective	In-patients	CIDI, structured interview	Diagnosis at index vs recall of MDD at Time 2	52
Kendler et al. (2001),[a]	1872	19 months	Retrospective	Twins	SCID	LTH of MDD at index vs Time 2	59
Thompson et al. (2004),[b]	419	13 years	Retrospective	Community	DIS	LTH of dysphoria or anhedonia at index vs Time 2	38
Wells and Horwood (2004)	374	4–10 years	Prospective	Community	DIS, CIDI	Recall at age 25 of LTH of anhedonia or dysphoria before age 21	44
Patten et al. (2012)	1688	2–10 years	Prospective	Community	CIDI-SFMD, additional probes	LTH of dysphoria or anhedonia or diagnosis of depression	41
Takayanagi et al. (2014),[b]	140	12–24 years	Retrospective	Community	DIS	LTH of MDD reported Times 1–3 vs LTH reported at Time 4	33

CIDI: Composite International Diagnostic Interview, CIDI-SMFD: Composite International Diagnostic Interview- Short Form for Major Depression, DIS: Diagnostic Interview Schedule, RDC: Research Diagnostic Criteria, SADS-L: Schedule for Affective Disorders and Schizophrenia-Lifetime, SCID: Structured Clinical Interview for DSM-IV.

[a] These two studies evaluated data from the Virginia Twin Registry.

[b] These two studies evaluated data from the Baltimore Epidemiologic Catchment Area Survey.

because it encompasses a large amount of variability in severity of episodes, which could influence recall (Warshaw et al., 1991).

Both prospective and retrospective methods of assessing recall accuracy have been utilized. Retrospective studies are more numerous, and typically involved asking patients about past MDEs, then attempting to match those recollections to previously documented clinical notes. Prospective approaches are more rigorous, with patients' symptoms specifically assessed at one time point, followed by asking patients to recall those symptoms at a later time-point. Overall, prospective monitoring for MDEs found rates roughly double those identified by retrospective inquiry (Moffitt et al., 2010).

In the current analysis, we aimed to assess prospectively the accuracy of patients' recall of an index MDE among adult outpatients participating in a 2-year clinical trial. We hypothesized that recall of sufficient symptoms to identify the index MDE would be higher than that observed in community studies in which contact with clinicians is less frequent. A second aim was to investigate the accuracy of recall of specific depressive symptoms, which has not previously been examined. We hypothesized that more salient symptoms, such as suicidal ideation, would be more accurately recalled than less depression-specific symptoms, such as impaired concentration.

2. Material and methods

2.1. Study design

The Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study protocol has been published previously (Dunlop et al., 2012). The study randomized 344 adults with MDD to receive 12 weeks of treatment with either cognitive behavior therapy (CBT, 16 1-h individual sessions), duloxetine (30–60 mg/day) or escitalopram (10–20 mg/day). Remission was defined as a score ≤ 7 at both weeks 10 and 12 on the 17-item Hamilton Depressive Rating Scale (HDRS-17, Hamilton, 1960). Patients who remitted at the end of 12 weeks were offered participation in a long-term follow-up study lasting 21 months. Those that did not remit were eligible for an additional 12 weeks of treatment, during which CBT was added to medication non-remitters, and escitalopram added to CBT non-remitters. At the end of week 24, patients in the combination treatment who demonstrated $\geq 50\%$ reduction from their baseline HDRS-17 score were offered the same long-term follow-up protocol as the monotherapy remitters for a duration of 18 months. During long-term follow-up, patients completed visits comprised of a clinical interview with a psychiatrist, clinician ratings and self-report measures every 3 months for the duration of the follow-up period, or until they met criteria for recurrence. Patients receiving CBT received three booster sessions and could attend one "crisis" session during each of the two years of long-term follow-up. Patients taking an antidepressant remained on their medication, with the option of tapering off the medicine after month 12. Participation in the study was terminated if the patient was lost to follow up, experienced a depressive recurrence, or completed the full two years of the study. Clinical outcomes from monotherapy treatment phase (Dunlop et al., 2017), the combination treatment (Dunlop et al., 2019), and long-term follow-up phase (Kennedy et al., 2018) are published elsewhere.

2.2. Participants

Treatment-naïve adults (18–65 years old) fluent in English or Spanish who had a primary psychiatric diagnosis of moderate-to-severe MDD without psychotic features were recruited from the metropolitan Atlanta area. Study participants were required to have a primary psychiatric diagnosis of MDD, as determined by the patient version of the Structured Clinical Interview for DSM-IV (SCID-IV) (First et al., 1995) and confirmed by a psychiatrist's interview. Eligibility also required a HDRS-17 score ≥ 18 at the screening visit and ≥ 15 at the

randomization (baseline) visit. In addition to possessing the ability to understand and provide informed consent, patients were required to have never previously received an evidence-based treatment for depression. This criterion was operationalized as never having taken an antidepressant at a minimum effective dose for ≥4 weeks, or having received an evidence-based psychotherapy for depression for ≥4 sessions. Exclusionary criteria included pregnancy or breastfeeding, unstable medical conditions or medical conditions that could interfere with interpretation of the study, a current or lifetime diagnosis of bipolar, psychotic, or cognitive disorders, a diagnosis of substance dependence, obsessive compulsive disorder, or an eating disorder in the past year, or substance abuse in the previous three months.

The researchers made the decision to assess patients' MDE recall accuracy after the study had been underway for five years, so recall data for this analysis were available for approximately the last third of the patients enrolled in PReDICT.

2.3. Assessment of MDE recall

The index MDE and presence of specific symptoms was defined at the screening visit via administration of the SCID-IV. Evaluation of recall accuracy was made by administering Module A (Mood disorders) of the SCID-IV at the months 12 and 24 follow-up visits. Trained research coordinators who administered Module A of the SCID-IV at the follow-up visits were blind to the SCID-IV assessment conducted at screening. When the SCID-IV module A was administered at the follow-up visits, patients were asked to recall the symptoms they had been experiencing at the time they entered the study, and they were provided with the dates of the one-month period preceding their screening visit as a memory aid.

The SCID-IV facilitates DSM-IV Axis I diagnoses through a validated semi-structured interview with good interrater reliability (Lobbestael et al., 2011). For each criterion symptom for an MDE in the SCID-IV, patients are rated as "1" if the symptom does not meet criteria, "2" if the symptom is present but at a subthreshold level, and "3" if the symptom meets the full threshold to count toward the diagnosis. To qualify as having a MDE on the SCID-IV, patients must be scored "3" for either depressed mood (item 1, dysphoria) or markedly diminished interest or pleasure (item 2, anhedonia), as well as at least four additional symptoms for a period of two weeks. Other symptoms addressed in the SCID-IV, with their respective item numbers, include: significant change in appetite or weight (3), insomnia or hypersomnia (4), psychomotor retardation or agitation (5), fatigue or loss of energy (6), feelings of worthlessness or excessive guilt (7), diminished ability to think, concentrate or make decisions (8), and thoughts of death or suicidal ideation (9) (First et al., 1995). Each symptom meeting criteria for MDE had to be endorsed by the patient as being present "most of the day, nearly every day," in the prior two weeks, except for thoughts of death or suicide, which only needed to occur more than once in the prior two weeks.

Patients were also assessed for depression severity at baseline and subsequent study visits using two clinician-rated and two self-report instruments. The Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Åsberg, 1979) and HDRS-17 were administered by raters blinded to treatment condition using structured interview guides for each instrument (Williams and Kobak, 2008; Williams, 1989). The MADRS assesses symptoms of the prior week and consists of ten items rated zero-six, with the tenth item inquiring about suicidal thoughts (Montgomery and Åsberg, 1979). The HDRS-17 also assesses past week depressive symptoms and consists of nine items rated 0–4 and eight items rated 0–2 (Hamilton, 1960). Question three of the HDRS-17 probes suicidality in the past week and is rated 0–4. Patients completed the Beck Depression Inventory (BDI-I) and the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR). The BDI-I, a 21-item questionnaire, requires patients to choose one of several statements pertaining to the severity of specific symptoms which best describes how they have felt in the past two weeks, rated 0–3 (Beck et al., 1961). Item nine of the BDI-I assesses thoughts of self-harm. The QIDS-SR consists of 17 items rated 0–3 that align with the DSM-IV MDD criteria and asks patients to select a response for each item that best describes how they felt over the previous week; item 12 assesses thoughts of death and suicide (Rush et al., 2003).

2.4. Statistical analysis

Clinical and demographic characteristics of patients who provided recall assessment data versus those who did not were analyzed to see if the recall group was representative of the larger sample. For these comparisons, independent sample t-tests were used for continuous variables and Chi-square tests for categorical data.

Percent agreement was used as the primary measure of MDE recall accuracy (calculated separately for month 12 and for month 24). This measure is the proportion of patients who accurately recalled a sufficient number and type of SCID-IV MDE symptoms from their index episode to meet criteria for an MDE.

Our second aim was to evaluate the accuracy with which each patient recalled each of the 9 specific depression symptoms. Patients' responses at the follow-up visit regarding whether they had been experiencing each MDE symptom at the study intake assessment were classified as accurate (either both "yes" or both "no" at the screening and follow-up time point) or inaccurate. Percent agreement was used as the primary measure of symptom recall accuracy.

Those who were inaccurate in recall of specific symptoms could have erred by "forgetting" they had been experiencing the symptom at intake (i.e., "yes" at intake and "no" when recalling at follow-up), or "over-endorsing" (i.e., "no" at intake and "yes" when recalling at follow-up). To evaluate whether errors in recall accuracy were due more to forgetting or over-endorsing, we used the McNemar mid-*p* test (Fagerland et al., 2013). The McNemar test is a commonly used for test for paired binary data because it makes allowance for the dependency in the data arising from the matching. The exact version of the McNemar test has been recommended for small sample sizes, but it suffers from being very conservative and results in unnecessarily large p-values. Like the exact test, the mid-p test does not exceed the nominal significance level, and it does so without the severe conservatism of the exact test (Fagerland et al., 2013). Due to the exploratory nature of this analysis, we did not control for multiple comparisons. All calculations were performed using IBM SPSS v.24.

3. Results

Of the 344 participants randomized in PReDICT, 79 completed MDE recall data for the month 12 or month 24 follow-up time-points (the "recall sample"). The remaining 265 randomized patients did not have recall data, either due to not entering the long-term follow-up phase or due to completing the long-term follow-up phase before the recall assessments were instituted. Of the 79 patients with recall data, 42 completed assessments at both the month 12 and month 24 visits. Eighteen participants were unique to the month 12 group (total n = 60), and 19 were unique to the month 24 group (total n = 61). At the recall assessment visits, nearly all patients were euthymic, with only three patients (two at month 12 and one at month 24) experiencing a recurrent MDE at the time of the recall assessment.

As shown in Table 2, other than the recall sample being more significantly likely to have recurrent depression, there were no significant differences in demographic or clinical variables between the recall (n = 79) and non-recall (n = 265) subsets of the PReDICT participants. Of the whole sample, 162 (47%) had recurrent MDD; among these patients the mean number of lifetime episodes was 3.67, excluding the 18 subjects who reported too many episodes to count. The index MDE was chronic in 106 (30.8%) of the whole sample; these patients had a mean episode duration of 301 ± 351 weeks. One or more current

Table 2
Clinical and demographic characteristics of participants who did and did not contribute MDE recall data.

	Non-Recall Sample (n = 265)		Recall Sample (n = 79)		χ2	p
	N	%	N	%		
Race						
White	126	47.5	38	48.1		
Black	47	17.7	17	21.5	1.93	.93
Native American	60	22.6	14	17.7		
Other	32	12.1	10	12.7		
Sex						
Female	148	55.8	48	60.8	.60	.44
Previous Suicide Attempt						
Yes	23	8.7	2	2.5	3.41	.07
Chronic Episode						
Yes	87	32.8	19	24.1	2.62	.11
≥High school education						
Yes	231	87.2	74	93.7	2.56	.11
Recurrent MDD						
Yes	116	43.8	46	58.2	4.50	.03
Comorbid Anxiety Disorder						
Yes	106	40	33	41.8	.08	.78
Treatment						
CBT	85	32.1	25	31.6	.59	.75
Escitalopram	90	34	29	36.7		
Duloxetine	90	34	25	31.6		
	Mean	SD	Mean	SD	t	p
Age	40.13	11.72	39.39	11.48	.50	.62
Age of onset	31.12	14.33	29.35	13.70	.97	.34
Episode length (wks)	124.5	242.3	100.9	198.1	.79	.43
Hamilton Depression Rating						
Baseline	19.99	3.73	19.06	3.78	1.94	.05
Month 12[a]	3.71	3.91	4.37	4.60	0.96	.34
Month 24[b]	2.18	2.60	3.33	3.71	1.91	.06
Hamilton Anxiety Rating						
Baseline	16.27	5.34	15.35	4.62	1.37	.17
Month 12[a]	3.35	3.08	5.07	5.19	2.32	.023
Month 24[b]	2.00	2.35	3.84	3.51	3.32	.001

[a] Based on n = 97 in non-recall sample and n = 60 in recall sample at month 12.
[b] Based on n = 55 in non-recall sample and n = 61 in recall sample at month 24.

comorbid anxiety disorders were present in 139 (40.4%) of the whole sample and did not significantly differ in frequency between the recall and non-recall groups. The proportion of patients with each specific current anxiety disorder in descending order of frequency was: specific phobia (15.1%), social phobia (14.8%), generalized anxiety disorder (12.8%), posttraumatic stress disorder (7.0%), panic disorder (2.9%), and anxiety disorder not otherwise specified (1.7%).

3.1. Accuracy of MDE recall

At month 12, 57/60 (95%) patients endorsed a sufficient number of symptoms to accurately recall their index MDE. At month 24, 52/61 (85.2%) patients accurately recalled their index MDE. In the month 12 group, there was no significant difference between the number of symptoms reported at screening (M = 7.07, SD = 1.13) and number recalled at month 12 (M = 7.02, SD = 1.69), t(59) = 0.224, p = .824. The month 24 group also had no significant difference between the number of symptoms reported at screening (M = 7.16, SD = 1.24) and recalled at month 24 (M = 6.75, SD = 2.01), t(60) = 1.36, p = .180.

Recall accuracy for the index MDE was non-significantly better at both recall assessments among patients who had received CBT in phase 1 (CBT monotherapy) or phase 2 (CBT-medication combination) compared to patients who remitted during phase 1 with medication alone (i.e., never received CBT). Failure to recall at Month 12 was 0/35 for CBT-treated versus 3/25 (12%) in medication-treated patients. At month 24, the recall failure rates were 3/32 (9.4%) for CBT-treated versus 6/20 (30%) in medication-treated patients.

3.2. Symptom agreement between baseline and follow-up

As shown in Table 3, at month 12 patients most accurately recalled previously endorsing anhedonia (95%) and dysphoria (95%) at screening. In contrast, patients struggled to remember correctly if they had experienced changes in psychomotor function, with 11/60 (18.3%) denying its presence during the index episode, and 16/60 (26.7%) endorsing having had psychomotor dysfunction at month 12, though denying it at screening. Across the nine MDE symptoms, only item 9 was significantly different in terms of forgetting versus over-endorsing during recall of the index MDE. Specifically, 12 patients at month 12 endorsed that they had had thoughts of death or suicide at screening, despite denying that symptom during the intake assessment; only two appeared to forget that they had experienced it ($p < .007$).

Like month 12, the most accurately recalled symptoms at month 24 were dysphoria (90.2%) and anhedonia (90.2%). At this visit, patients had the most difficulty with correctly remembering whether they had experienced appetite/weight change as part of their index episode, with 18/61 (30%) forgetting this symptom despite endorsing it at screening, and 11 others (18%) endorsing having had the symptom, though they had denied it at screening. Psychomotor function was also poorly recalled, with 44% (27/61) giving discordant responses. The results for recall of thoughts of death or suicide were similar to those for month 12, though not statistically significant, with more patients erroneously endorsing these thoughts at follow-up when they had been absent at intake rather than denying them when they been present.

3.3. Exploratory analysis of suicidal thoughts

We conducted further analyses of suicidal ideation using the depression symptom rating scales to explore why 16–20% of patients at the follow-up visits endorsed having experienced thoughts of death or suicide during their index episode though they had denied it at the initial evaluation. Among the month 12 group, 4/12 (33%) of the "over-endorsing" patients had responded affirmatively to suicidal thoughts on the clinician-rated scales (HDRS-17 or MADRS). In contrast, 8/12 (67%) had endorsed the suicide item on a self-report measure, most frequently on the BDI-I. Four patients (33%) had not endorsed suicidal ideation on any of the four rating scales. Every patient who scored positively (> 0) on the suicidal item of a clinician-administered scale also scored positively on a self-report measure, but 5/12 (42%) who reported suicidality on a self-report measure did not report the symptom on the clinician-administered scales nor the SCID-IV.

Among the month 24 group, 5/10 (50%) of the "over-endorsing" patients had scored positive for suicidal ideation on a clinician rating scale, and 50% had indicated suicidal thoughts on a self-report measure. Only one patient had indicated suicidal thoughts on a self-report but not on a clinician rating, and two patients had revealed suicidal ideation on clinician assessments but not on a self-report measure. Three (30%) of the patients at month 24 did not indicate suicidal thoughts or behaviors on any of the measures. Overall, 70% of the patients who responded negatively regarding having thoughts of death or suicide during the screening visit SCID-IV assessment had actually endorsed suicidal ideation on at least one depression severity measure.

A comparison of the patients who were concordant in their recall of the SCID-IV item 9 at the follow-up visits with those who discordantly over-endorsed the symptom found no significant differences in the demographic or clinical characteristics listed in Table 2.

Table 3
Symptom agreement between screening, month 12, and month 24 visits.

Symptom		Mo 12 (n = 60)		p-value	% Agree	Month 24 (n = 61)		p-value	% Agree
		Yes	No			Yes	No		
Dysphoria				.625	95.0			.015	90.2
Screening	Yes	56	2			54	6		
	No	1	1			0	1		
Anhedonia				–	95.0			–	90.2
Screening	Yes	57	3			55	6		
	No	0	0			0	0		
Appetite				.424	60.0			.201	52.5
Screening	Yes	31	14			29	18		
	No	10	5			11	3		
Sleep				.727	88.3			.092	80.3
Screening	Yes	51	3			47	9		
	No	4	2			3	2		
Psychomotor				.345	55.0			.185	55.7
Screening	Yes	17	11			23	10		
	No	16	16			17	11		
Fatigue				–	90.0			.039	86.9
Screening	Yes	54	6			53	7		
	No	0	0			1	0		
Guilt				.210	75.0			.360	70.5
Screening	Yes	42	10			40	11		
	No	5	3			7	3		
Concentration				.804	73.3			.839	62.3
Screening	Yes	39	8			36	12		
	No	8	5			11	2		
Suicide				.007[a]	76.7			.332	73.8
Screening	Yes	18	2			15	6		
	No	12	28			10	30		

[a] $p < .05$.

4. Discussion

In this report of patients followed for up to 2 years as part of a clinical trial for MDD, we found patients' accuracy of recall for their index MDE to be higher than that reported in previous community-based studies. Specifically, one year after their initial assessment, 95% of patients recalled enough symptoms to identify accurately their index MDE, and at two years, 85% had accurate recall. Patients were particularly accurate in recalling both depressed mood and anhedonia symptoms, with more than 90% at both the 1- and 2-year follow-up time points accurately recalling these symptoms.

Our hypothesis that more severe symptoms, such as suicidal behavior, would be more accurately recalled than less salient symptoms was not confirmed, with 16–20% of the patients reporting having had suicidal thoughts during their index MDE when questioned at follow-up, even though they had denied the symptom at their original assessment. Prior research indicates that patients who experienced suicidal ideation as part of prior MDEs are substantially more likely to recall key aspects of their depressive episodes than patients who did not experience suicidal thoughts (Wells and Horwood, 2004). Thus, it is doubtful that patients were truly in error in recalling whether they had experienced such thoughts; rather, by the time of the month 12 and 24 assessments, they likely felt sufficiently comfortable with the research team to reveal thoughts they had kept hidden at the beginning of the study (Yigletu et al., 2004). Taken together, these outcomes indicate that patients in continuous follow-up care may be considered to have good recall of depressive symptoms up to two years prior, but that patient reluctance to report thoughts of death or suicide at intake remains an important clinical issue.

Our higher rate of recall of the index MDE compared to prior studies could derive from several factors. Most importantly, the repeated assessment of symptoms via depression rating scales (ranging from 14 to 27 assessments, depending on the treatment received and duration of follow-up) may have provided psychoeducational effects. These evaluations may have also increased memory for the specific symptoms of MDD, rather than more global impressions that may be formed in patients not undergoing repeated assessments. Furthermore, several community studies demonstrated that treatment-seeking patients have more accurate recall than non-treatment-seeking individuals with MDD (Kendler et al., 2001; Wells and Horwood, 2004).

Among patients with discordant recall of specific MDE symptoms, failure to recall specific symptoms originally endorsed at intake was a more common error than newly endorsing a symptom that was originally denied. Only for the symptoms of psychomotor change and suicidal ideation was the error of new endorsement more common than forgetting; this pattern was present at both the month 12 and month 24 assessments. Symptom recall failure may relate to mood congruency; some data suggest that patients more accurately recall past depressive symptoms if they are depressed at the time of assessing recall (Aneshensel et al., 1987; Kendler et al., 2001; Thompson et al., 2004). We could not explore this possibility because only three patients in our dataset were experiencing a depressive recurrence at the time of recall assessment. Interestingly, the converse of this evaluation (i.e., accuracy of recall for well periods among people in a current MDE) does not seem to have been evaluated in the literature, though it also can have clinical relevance. Future prospective research of these questions could incorporate ecological momentary assessments, which would provide denser datasets to evaluate the effects of time, mood, and life events on symptom recall accuracy (Ebner-Priemer and Trull, 2009).

Our results also demonstrated a discrepancy between reporting during clinician interviews and on self-report measures. Seventy-five percent of month 12 patients and 45% of month 24 patients who did not endorse suicidal ideation on the SCID-IV reported such thoughts on a self-report scale. Nearly all patients that reported thoughts of death or suicide on a self-report scale also did so on a clinician-rating scale. These results are consistent with prior studies finding that participants were more likely to endorse suicidal thoughts on a self-report measure than clinician interview (Kaplan et al., 1994; Yigletu et al., 2004; Gao et al., 2015). Underreporting suicidal thoughts at initial assessment may be attributed to several possible factors, including fear of

consequences from revealing suicidal thoughts, such as stigmatization, involuntary hospitalization, or exclusion from study participation (Kučukalić and Kučukalić, 2017). Repeated exposure to symptom terminology via assessments might have led to awareness that their feelings and thoughts actually did constitute a form of suicidal ideation that they had previously dismissed (Kaplan et al., 1994; Yigletu et al., 2004).

It is worth considering how the rating scales differentially anchor suicidal thoughts from "0" (symptom absent) to "1" (symptom present). On the BDI-I, an item 9 score of "1" is linked to the response, "I have thoughts of killing myself, but I would not carry them out" (Beck et al., 1961). This description of active suicidal ideation differs from a "1" on the QIDS-SR item 12, which states, "I feel that life is empty or wonder if it's worth living," reflecting less severe, passive thoughts (Rush et al., 2003). On the MADRS, a score of "2" is anchored by feeling "weary of life" and having "only fleeting suicidal thoughts" (Montgomery and Åsberg, 1979). A "1" on this MADRS item reflects an intensity or frequency of thoughts lower than those warranting a "2." On the HAMD-17, the anchor for a score of "1" is that the patient "feels life is not worth living" (Hamilton, 1960). Despite the apparently greater severity of a score of "1" on the BDI-I compared to a "1" on the other scales, more patients endorsed suicidal thoughts on this questionnaire than the other three. The 2-week period covered by the BDI-I, versus the 1-week of the other scales, may have contributed to this finding.

Several studies in Table 1 caution that their results are limited by a lack of independent validation that a participant had truly previously met criteria for an MDE (Andrews et al., 1999; Aneshensel et al., 1987). A strength of our analysis is the utilization of a well-defined index MDE, enabling valid recall comparisons. Another strength was the ability to evaluate recall accuracy for specific symptoms, a novel addition to the literature.

The study was limited by the relatively small sample size compared to those used in community surveys. Prior research has reported that greater depression severity of the index episode increased the likelihood of accurately recalling a lifetime history of an MDE (Foley et al., 1998; Kendler et al., 2001; Wells and Horwood, 2004). Due to the low rates of recall failure, we were unable to perform analyses that could help identify predictors of recall accuracy. The generalizability of our results may be limited by our sample being treatment-naïve and having a high education level. In addition, patients with active substance use disorders were excluded and only 40% of our sample had a comorbid anxiety disorder, roughly a third lower than to the 57% reported in the National Comorbidity Survey Replication (Kessler et al., 2003). By nature of being treatment-naive, patients in PReDICT may have been particularly more reluctant to disclose suicidal thoughts at the screening assessment compared to patients with experience in psychiatric care settings, where questions about suicidal thoughts are routine. Over 96% of our sample was euthymic at the time of recall assessment, so the impact of mood congruency on recall accuracy could not be evaluated. Our 2-year period of follow-up of was relatively brief; longer periods of follow-up may have found lower rates of recall accuracy.

5. Conclusions

The great majority of patients who have participated in a clinical trial can reliably report a sufficient number of symptoms to identify a prior MDE occurring within the past 2 years. These results suggest that patients in clinical care are likely to have greater recall accuracy for past MDEs than the low accuracy identified in earlier studies of community samples involving many patients not in active treatment. However, recall of specific symptoms other than dysphoria and anhedonia is not as dependable. For thoughts of death or suicide, clinicians need be vigilant for under-reporting among patients with MDD, particularly among those newly presenting for treatment. Employing a multimodal approach with both self-report and clinician interviews more sensitively detects this important symptom.

Role of the funding source

This work was supported by the National Institutes of Health (H.S.M., P50 MH077083; W.E.C, RO1MH080880; B.W.D., K23 MH086690; and David S. Stephens, UL1 RR025008, M01 RR0039).

References

American Psychiatric Association, 2010. Practice Guideline for the Treatment of Patients with Major Depressive Disorder, third ed. American Psychiatric Association, Washington, D.C.

Andrews, G., Anstey, K., Brodaty, H., Issakidis, C., Luscombe, G., 1999. Recall of depressive episode 25 years previously. Psychol. Med. 29, 787–791.

Aneshensel, C.S., Estrada, A.L., Hansell, M.J., Clark, V.A., 1987. Social psychological aspects of reporting behavior: lifetime depressive episode reports. J. Health Soc. Beyond Behav. 28, 232–246.

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

Burcusa, S.L., Iacono, W.G., 2007. Risk for recurrence in depression. Clin. Psychol. Rev. 27, 959–985.

Dunlop, B.W., Binder, E.B., Cubells, J.F., Goodman, M.G., Kelley, M.E., Kinkead, B., Kutnerm, M., Nemeroff, C.B., Newport, D.J., Owens, M.J., Pace, T.W.W., Ritchie, J.C., Aponte-Rivera, V., Westen, D., Craighead, W.E., Mayberg, H.S., 2012. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. Trials 13, 106.

Dunlop, B.W., Kelley, M.E., Aponte-Rivera, V., Kinkead, B., Mletzko-Crowe, T., Ritchie, J.C., Nemeroff, C.B., Craighead, W.E., Mayberg, H.S., 2017. Effects of patient preferences on outcomes in the predictors of remission in depression to individual and combined treatments (PReDICT) study. Am. J. Psychiatry 174, 546–556.

Dunlop, B.W., LoParo, D., Kinkead, B., Mletzko-Crowe, T., Cole, S.P., Nemeroff, C.B., Mayberg, H.S., Craighead, W.E., 2019 Feb 15. Benefits of sequentially adding cognitive-behavioral therapy or antidepressant medication for adults With nonremitting depression. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.2018.18091075. [Epub ahead of print], appiajp201818091075.

Ebner-Priemer, U.W., Trull, T.J., 2009. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol. Assess. 21, 463–475.

Fagerland, M.W., Lydersen, S., Laake, P., 2013. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC Med. Res. Methodol. 13, 91.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1995. Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID-I/P, Version 2.0). Biometrics Research Department, New York State Psychiatric Institute, New York.

Foley, D., Neale, M., Kendler, K., 1998. Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. Psychol. Med. 28, 857–870.

Gao, K., Wu, R., Wang, Z., Ren, M., Kemp, D.E., Chan, P.K., Conroy, C.M., Serrano, M.B., Ganocy, S.J., Calabrese, J.R., 2015. Disagreement between self-reported and clinician-ascertained suicidal ideation and its correlation with depression and anxiety severity in patients with major depressive disorder or bipolar disorder. J. Psychiatr. Res. 60, 117–124.

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

Kaplan, M.L., Asnis, G.M., Sanderson, W.C., Keswani, L., De Lecuona, J.M., Joseph, S., 1994. Suicide assessment: clinical interview vs. self-report. J. Clin. Psychol. 50, 294–298.

Kendler, K., Gardner, C., Prescott, C., 2001. Are there sex differences in the reliability of a lifetime history of major depression and its predictors? Psychol. Med. 31, 617–625.

Kennedy, J.C., Dunlop, B.W., Craighead, L.W., Nemeroff, C.B., Mayberg, H.S., Craighead, W.E., 2018. Follow-up of monotherapy remitters in the PReDICT study: maintenance treatment outcomes and clinical predictors of recurrence. J. Consult. Clin. Psychiatry 86, 189–199.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 289, 3095–3105.

Kučukalić, S., Kučukalić, A., 2017. Stigma and suicide. Psychiatr. Danub. 29, 895–899.

Lobbestael, J., Leurgans, M., Arntz, A., 2011. Inter-rater reliability of the structured clinical interview for DSM-IV Axis I disorders (SCID I) and Axis II disorders (SCID II). Clin. Psychol. Psychother. 18, 75–79.

Moffitt, T., Caspi, A., Taylor, A., Kokaua, J., Milne, B., Polanczyk, G., Poulton, R., 2010. How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. Psychol. Med. 40, 899–909.

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

Patten, S.B., Williams, J.A., Lavorato, D.H., Bulloch, A.M., D'Arcy, C., Streiner, D.L., 2012. Recall of recent and more remote depressive episodes in a prospective cohort study. Soc. Psychiatr. Psychiatr. Epidemiol. 47, 691–696.

Rohde, P., Lewinsohn, P.M., Klein, D.N., Seeley, J.R., Gau, J.M., 2013. Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, and adulthood. Clin. Psychol. Sci. 1, 41–53.

Rush, A.J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Biggs, M.M., Suppes, T., Crismon, M.L., Shores-Wilson, K., Toprac, M.G., Dennehy, E.B., Witte, B., Kashner, T.M., 2003. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating

(QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psychiatry 54, 573–583.
Solomon, D.A., Keller, M.B., Leon, A.C., Mueller, T.I., Lavori, P.W., Shea, M.T., Coryell, W., Warshaw, M., Turvey, C., Maser, J.D., Endicott, J., 2000. Multiple recurrences of major depressive disorder. Am. J. Psychiatry 157, 229–233.
Takayanagi, Y., Spira, A.P., Roth, K.B., Gallo, J.J., Eaton, W.W., Mojtabai, R., 2014. Accuracy of reports of lifetime mental and physical disorders: results from the Baltimore Epidemiological Catchment Area Study. JAMA Psychiatry 71, 273–280.
Thompson, R., Bogner, H.R., Coyne, J.C., Gallo, J.J., Eaton, W.W., 2004. Personal characteristics associated with consistency of recall of depressed or anhedonic mood in the 13-year follow-up of the Baltimore Epidemiologic Catchment Area Survey. Acta Psychiatr. Scand. 109, 345–354.
Warshaw, M.G., Klerman, G.L., Lavori, P.W., 1991. Are secular trends in major depression an artifact of recall? J. Psychiatr. Res. 25, 141–151.
Wells, J.E., Horwood, L.J., 2004. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol. Med. 34, 1001–1011.
Williams, J.B., 1989. A structured interview guide for the Hamilton Depression Rating. Arch. Gen. Psychiatr. 45, 742–747 (Spanish translation, Jan 2009, obtained from author).
Williams, J.B.W., Kobak, K.A., 2008. Development and reliability of the SIGMA: a structured interview guide for the montgomery-asberg depression rating scale (MADRS). Br. J. Psychiatry 192, 52–58 (Spanish translation, Jan 2009, obtained from author).
World Health Organization, 2017. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, Geneva Licence: CC BY-NC-SA 3.0 IGO.
Yigletu, H., Tucker, S., Harris, M., Hatlevig, J., 2004. Assessing suicide ideation: comparing self-report versus clinician report. J. Am. Psychiatr. Nurses Assoc. 10, 9–15.