Psychiatry Research 292 (2020) 113313

Contents lists available at ScienceDirect

Psychiatry Research

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

Test-retest & familial concordance of MDD symptoms

Ariela J.E. Kaiser[a,b], Carter J. Funkhouser[a,b], Vijay A. Mittal[b,c], Sebastian Walther[d], Stewart A. Shankman[a,b,*]

[a] University of Illinois at Chicago, Department of Psychology, United States
[b] Northwestern University, Department of Psychiatry and Behavioral Sciences, United States
[c] Northwestern University, Departments of Psychology, Medical Social Sciences.. Institutes for Policy Research, Innovations in Developmental Sciences (DevSci), United States
[d] University of Bern, University Hospital of Psychiatry, Translational Research Center, Bern, Switzerland

ARTICLE INFO

Keywords:
Depression
Psychometrics
Symptoms
Heterogeneity

ABSTRACT

Psychopathology research has increasingly sought to study the etiology and treatment of individual symptoms, rather than categorical diagnoses. However, it is unclear whether commonly used measures have adequate psychometric properties for assessing individual symptoms. This study examined the test-retest reliability and familial concordance (an indicator of validity) of the symptoms of Major Depressive Disorder (MDD), a disorder consisting of nine core symptoms, most of which are aggregated (e.g., symptom 7 of the DSM criteria for MDD is worthlessness *or* guilt). Lifetime MDD symptoms were measured in 504 young adults (237 sibling pairs) using the Structured Clinical Interview for DSM-5 (SCID). Fifty-one people completed a second SCID within three weeks of their first SCID. Results indicated that aggregated and unaggregated symptoms demonstrated moderate to substantial test-retest reliability and generally significant, but slight to fair familial concordance (with the highest familial concordance being for markedly diminished interest or pleasure and its unaggregated components – decreased interest and decreased pleasure). Given the increasing focus on the differential validity of individual MDD symptoms, the present study suggests that interview-based assessments of depression can assess most individual symptoms with adequate levels of reliability and validity.

1. Introduction

Major Depressive Disorder (MDD) is one of the most common psychiatric disorders, is a major cause of disability worldwide, and is among the leading disorders for global disease burden (Lopez et al., 2006). MDD often co-occurs with other psychiatric disorders (Kessler et al., 2005; Lewinsohn et al., 2004), is a significant predictor of suicide (Berman, 2009), and is highly recurring (e.g., more than half of all individuals diagnosed with MDD experience more than one episode in their lifetime [McClintock et al., 2010]).

MDD, like most psychiatric disorders, is a polythetic disorder, which means the disorder is defined by multiple symptoms, and not all symptoms need to be present in order for the syndrome to be present. In the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), individuals meet criteria for a major depressive episode if they endorse five of nine symptoms, at least one of which must be either depressed mood or markedly diminished interest/pleasure. An underlying assumption of the operationalization of MDD as a categorical diagnosis is that the non-cardinal symptoms are interchangeable. In other words, the number of symptoms is emphasized rather than the nature of the symptoms.

The MDD categorical diagnosis has also shown questionable reliability across multiple studies (Markon et al., 2011; Bromet et al., 1986; Spitzer and Fleiss, 1974). In the field trials for DSM-5, clinician agreement for the MDD diagnosis was 0.28 (95% CI 0.20–0.35), which fell into in the questionable range of pooled intraclass Kappa (Regier et al., 2013). The poor reliability also likely contributes to difficulty clarifying the etiology of MDD. That is, if an assessment of a construct or disorder has poor reliability, then it will be extremely difficult to identify its etiology (Smoller and Finn, 2003).

One explanation for the poor reliability of the MDD diagnosis may be the phenotypic heterogeneity of depression. There are thousands of possible symptom presentations that meet diagnostic criteria for MDD (Zimmerman et al., 2015) – for example, one study identified 1,030 unique depression symptom profiles in 3,703 individuals with MDD (Fried and Nesse, 2015). What further compounds the problem of the polythetic criteria is that 7 of the 9 symptoms are aggregated, consisting of at least two different symptoms (e.g., worthlessness OR guilt).

* Corresponding author at: Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL, 60611, United States
E-mail address: stew.shankman@northwestern.edu (S.A. Shankman).

https://doi.org/10.1016/j.psychres.2020.113313
Received 29 April 2020; Received in revised form 15 July 2020; Accepted 19 July 2020
Available online 20 July 2020
0165-1781/

Moreover, many of the symptoms consist of contrasting features (e.g., increase OR decrease in weight/appetite). Importantly, studies show that many of the aggregated symptoms that are phrased as contrasting co-occur in some patients with MDD (e.g., psychomotor retardation OR psychomotor agitation (Parker et al., 1995); insomnia OR hypersomnia Soehner et al., 2014).

Although the categorical diagnosis has demonstrated questionable reliability, it remains unclear if the individual symptoms comprising MDD are themselves reliable. One of the main challenges in symptom-based research is establishing reliable and valid tools to measure individual symptoms. Although most rating scales are not validated at the symptom level (and certainly not for unaggregated symptoms), one small study (N = 31) did examine inter-rater reliability and concurrent validity for individual symptoms of MDD (Mazure et al., 1986). Results highlighted that most symptoms of MDD can be reliably detected by clinicians during a semi-structured interview and that observable symptoms correlated with patients' behaviors.

1.1. Differential validity of individual depressive symptoms

In addition to reliability, research has increasingly focused on the differential validity of individual depressive symptoms. For example, studies suggest that diverse etiological factors may lead to the occurrence of different depressive symptoms, further suggesting that different depressive symptoms are not interchangeable. Keller et al. (2007) found that chronic stress was associated with some but not all symptoms, such as fatigue and hypersomnia. Likewise, genetic disposition to specific symptoms is uncertain, as Kendler and Aggen (2017) detected only modest concordance of individual MDD symptoms among monozygotic twins, while there were separate environmental factors for different symptoms. Another study found that chronic stress, prior history of depression, higher number of stressful life events, childhood stress, and sex predicted worsening of only certain MDD symptoms (Fried et al., 2014). Finally, a subset of MDD symptoms seems to impact functional impairment, e.g. low mood, difficulty concentrating, fatigue, and loss of interest/pleasure (Fried and Nesse, 2014). These findings support the idea that symptom sum-scores disregard qualitative differences between particular depressive symptoms.

Besides research showing differential impact of certain MDD symptoms on impairment and etiological factors, several studies have shown that the presence of particular symptoms (e.g., sleep disturbances and hopelessness [Benca and Peterson, 2008; Dew et al., 1997]) predicted a worse outcome to antidepressant treatment (see also, McMakin et al., 2012; Uher et al.,2012). Thus, treating specific symptoms as interchangeable indicators of a latent disease ignores key prognostic variables in the course and treatment of depression.

Studying specific symptoms, rather than a latent disorder, is also consistent with the network theory of psychopathology, which states that disorders are the result of causal relationships between individual symptoms (Borsboom and Cramer, 2013; Boccaletti et al., 2006). Psychopathology network analysis has become extremely popular in psychopathology research (Funkhouser et al., 2020; McNally, 2019) suggesting that even more MDD research will focus on individual symptoms. Additionally, studying the psychometrics of individual symptoms would support the growing interest in depression measures guided by Item Response Theory, such as computerized adaptive tests that automatically administer items to examinees based on their prior responses (Fries et al., 2011).

In sum, individual symptoms of MDD have been and continue to be widely studied, and the exponentially increasing popularity of network theory and modeling suggests that even more MDD research will focus on individual symptoms. Despite the increasing focus on individual symptoms in MDD research, relatively little is known about the reliability of commonly used measures of individual MDD symptoms.

The goal of the study was to examine the psychometric properties for the (1) categorical diagnosis, (2) aggregated symptoms (e.g., sleep disturbance overall), and (3) unaggregated symptoms (e.g., hypersomnia and insomnia) of MDD. The study focused on two psychometric properties - test-retest reliability and familialness (i.e., the extent to which they run in families). Since Robins and Guze's (1970) classic paper on psychiatric validity, familialness has long been considered an important validator for psychiatric disorders (and, by extension, individual symptoms) given that family history of a disorder is a robust correlate/risk factor for disorders in probands (Klein et al., 2003; Nierenberg et al., 2007). To our knowledge, no study has examined the test-retest reliability of all unaggregated and aggregated symptoms, and only one study examined the familialness of individual symptoms. Korszun et al. (2004) examined the familialness of depression symptom dimensions between sibling pairs and found the highest correlations between siblings for the symptoms of restlessness, anxiety, loss of libido, and irritability (rs = .26-.31; Korszun, et al., 2004). The present study will build off Korszun et al. by examining DSM-5 aggregated and unaggregated symptoms in a community sample of sibling pairs rather than in a sample with severe and recurrent MDD. This is an important advancement because individuals with subthreshold MDD are a clinically significant group (Shankman et al., 2008, 2009) and only focusing on those with severe and recurrent MDD may limit generalizability. Given the differential validity of individual MDD symptoms, certain symptoms may run in families more than others, further supporting the varied validity of individual MDD symptoms.

We hypothesize that both aggregated and unaggregated symptoms will exhibit significant test-retest reliability and familial concordance, but (given the heterogeneity of MDD symptoms) will vary substantially.

2. Methods

2.1. Participants

504 participants were recruited as part of a NIMH-funded family study (see Gorka et al., 2016for additional details). Participants were 18 to 30 years old, nested within 274 families, and included 237 sibling pairs breaking down into 40 male-male dyads, 91 male-female dyads, and 106 female-female dyads (see Table 1 for additional participant demographics). Advertisements (fliers, internet postings, etc.) were used to recruit participants from the community and from mental health clinics. A Research Domain Criteria (RDoC) approach (Shankman and Gorka, 2015) was taken to participant recruitment such that recruitment screening was agnostic to DSM diagnostic categories (beyond the exclusion criteria listed below). However, participants with elevated symptoms of internalizing psychopathology were oversampled to ensure that the sample was clinically relevant. Specifically, the Depression, Anxiety, and Stress Scale (Lovibond and Lovibond, 1995) was administered during the initial phone screen to ensure that the severity of internalizing symptomology within the sample was normally distributed, but also was higher than the general population (M = 10.35 [SD = 10.07] vs. M =8.3 [SD = 9.8]; Crawford et al., 2011).

Table 1
Participant demographics and characteristics.

	Sibling Pairs (N = 237)	Test-Retest Reliability (N = 51)
Age (Mean, SD)	22.3 (3.2)	22.3 (3.3)
Percent Female	64%	60%
Race/Ethnicity		
Percent Caucasian	41.1%	41.2%
Percent Hispanic	22.6%	11.8%
Percent African American	14.5%	23.5%
Percent Asian	10.6%	9.8%
Percent Middle eastern	4.1%	2.0%
Percent Mixed race	6.1%	9.0%
Percent Other	.9%	2.0%

Inclusion criteria specified that participants had at least one full biological sibling that was also willing to participate in the study. Exclusion criteria included personal or family history of psychosis or mania at the time of the interview (given that psychosis and mania have been shown to be separable from internalizing and externalizing disorders; Caspi et al., 2014; Kotov et al., 2017; Krueger et al., 1999, Markon, 2010), being a twin, inability to read or write in English, history of serious head trauma, and left-handedness (to protect against confounds with the neurophysiological data collected for other aims of the larger study).

2.2. Structured clinical interview for diagnostic and statistical manual of mental disorders (SCID)

Symptoms were assessed for participants' worst lifetime depressive episode using the SCID (First et al., 2015), a widely used semi-structured clinical interview. The SCID used in the present study was modified slightly to better assess individual symptoms (Shankman et al., 2018). First, the separate parts of aggregated symptoms were coded independently (e.g., the MDD symptom "worthlessness or guilt" was split so that it yielded separate ratings for worthlessness and guilt). Second, to increase sensitivity to individuals with subthreshold psychopathology and facilitate the calculation of symptom severity scales, interviewers assessed all lifetime MDD symptoms even if the cardinal symptoms (depressed mood and loss of interest/pleasure) were not endorsed (note: if neither cardinal symptom was endorsed, interviewers assessed symptoms for the 2-week or longer period in their life during which participants were "most distressed or upset"). Doctoral students and bachelor's level interviewers were supervised by a licensed clinical psychologist and trained to criterion on the SCID by viewing the SCID-101 training videos (SCID-101, 1998), overseeing two or three SCID interviews with an experienced interviewer, and completing three SCID interviews (observed by an advanced interviewer) in which diagnoses were in full agreement with those of the observer. For this study, three sets of variables were pulled from the adapted SCID: the categorical diagnosis of MDD, the nine aggregated symptoms of MDD, and the 23 unaggregated symptoms of MDD. To assess test-retest reliability, a subset of participants (N = 51) were pseudo-randomly selected from the overall sample to complete a second SCID with a different interviewer within three weeks of their first SCID (M = 8.5 days, SD = 4.3; see Table 1 for subsample characteristics). As SCID ratings for each of the depressive items consist of 1 (absent), 2 (subthreshold) and 3 (full threshold), ratings were dichotomized to reflect the presence (3) or absence (1 or 2) of each symptom. Similarly, diagnoses of lifetime MDD reflected the presence or absence of MDD, and the number of depressive episodes was not considered.

2.3. Data analyses

The test-retest reliability and familialness of the lifetime MDD diagnosis and each lifetime MDD symptom were evaluated using Cohen's kappa using established conventions for agreement (Cohen, 1960). We used R (Version 1.2.1237; R Core Team, 2019) and the R-packages psych (Version 1.8.12; Revelle, 2018) and boot (Version 1.2-20; Canty and Ripley, 2013). For power analyses, we used package irr (Version 0.84.1; Gamer et al., 2012). We randomly assigned each subject within each sibling pair to be either sibling 1 or sibling 2 and then computed bootstrapped 95% confidence intervals around the estimated Kappa for each symptom (i.e., the 9 aggregated symptoms and 23 unaggregated symptoms). If the confidence interval of the Kappa did not contain a value of zero, it was considered significant.

Sex and age have been shown to impact the clinical presentation of MDD. Women are approximately 1.7 times as likely as men to report a lifetime history of MDD (Kessler et al., 1993) and research has shown that the presentation of MDD symptoms varies as a function of age (Kovacs et al., 1997; Lux and Kendler, 2010). In order to include age and sex as covariates, logistic regressions were run as kappa analyses do not allow for the inclusion of covariates. Thus, all models testing for familial concordance covaried for each sibling's sex and age and the interaction of the two siblings' sex and age. A randomly assigned sibling's symptom (aggregated or unaggregated) was the independent variable, the other sibling's symptom was the dependent variable, and the sexes and ages of the two siblings and their interactions were included as covariates (for a similar approach, see Khan et al., 2002; Moskvina et al., 2008). If the same familial effect was operative for both male and female siblings, logistic regressions covarying for the interaction of each sibling's sex as well as their ages would remain significant. However, if a familial effect was age- and/or sex-dependent, then the results would no longer remain when including these covariates.

3. Results

3.1. Test-retest reliability

The lifetime diagnosis of MDD had substantial test-retest reliability (k = .68). The test-retest reliabilities for aggregated symptoms fell into the moderate range (see Fig. 1a), and were highest for loss of interest/pleasure (k = .71) and depressed mood (k = .63) and lowest for suicidal thoughts and behaviors (k = .52) and inappropriate guilt and feelings of worthlessness (k = .52). At the unaggregated symptom level, nearly all unaggregated symptoms had fair to substantial test-retest reliability (see Fig. 1b). The unaggregated symptoms with the strongest reliability (all of which fell into the substantial range) included middle insomnia (k = .75), hypersomnia (k = .72), psychomotor retardation (k = .72), and specific plan for a suicide attempt (k = .79). The unaggregated symptoms with the lowest test-retest reliability included weight loss (k = .17) and psychomotor agitation (k = .23) (Table 2).

3.2. Familial concordance

In the sample of 237 sibling pairs, there was fair familial concordance for the lifetime categorical MDD diagnosis (k = .21). The familial concordances for the aggregated symptoms are presented in Fig. 2 and were highest for loss of interest/pleasure (k = .28) and depressed mood (k = .21). At the unaggregated symptom level, the symptoms that comprise anhedonialoss of interest/pleasure —loss of interest (k = .21) and loss of pleasure (k = .25)—had the highest familial concordance and were in the fair agreement range (see Fig. 3). The unaggregated symptoms with the lowest familial concordance included sleep disturbance (k = .08) and psychomotor disturbance (k = .04).

3.3. Impact of sex and age on familial concordance

Results of logistic regressions examining familial associations covarying for age and sex were consistent with the results of Cohen's Kappa for all aggregated symptoms and the majority of unaggregated symptoms in MDD. However, the unaggregated symptoms of appetite loss (B = .80, p = .035) and suicidal ideation (B = .96, p = .025) were significant when controlling for sex and age but did not have significant Cohen's Kappas in models that did not take into account sex and age differences.

4. Discussion

Research has increasingly studied the etiology, course, and treatment responsiveness of individual MDD symptoms rather than the overall disorder. Many of these studies have assessed symptoms using single items from structured interviews. However, the psychometric properties of these assessments are unclear. Therefore, it is necessary to study the reliability and validity of tools that can measure individual

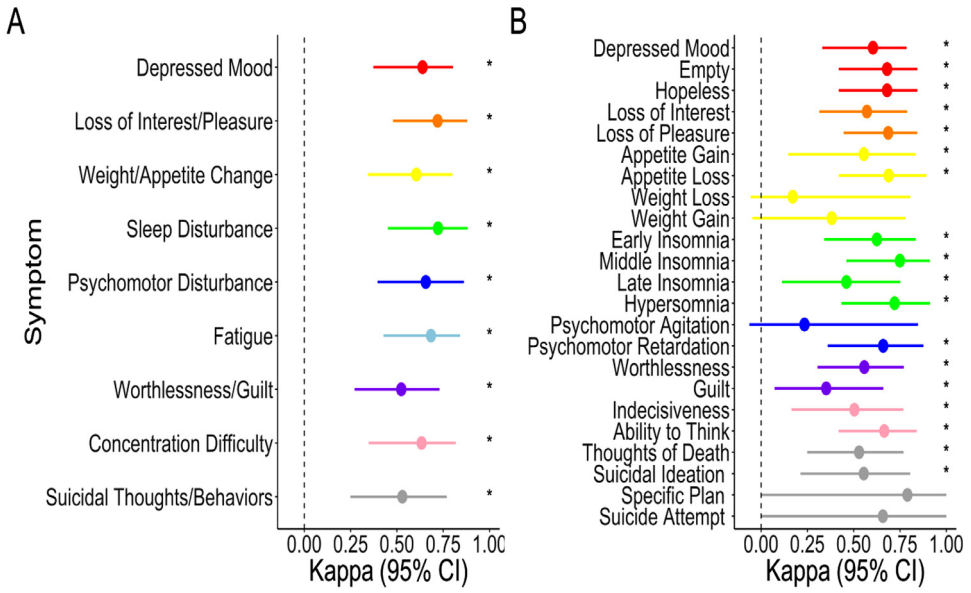

Fig. 1. Cohen's Kappas for the test-retest reliability of (a) aggregated and (b) unaggregated MDD symptoms.

Table 2
Aggregated and unaggregated symptom endorsement at Time 1, Time 2, both, only one time.

	% Endorsed Time 1	% Endorsed Time 2	% Endorsed Time 1 and Time 2	% Endorsed at one time point only (but not both)
Depressed Mood	56.86	58.82	49.02	17.65
Low Mood	54.9	54.9	45.1	19.61
Emptiness	43.14	43.14	35.29	15.69
Hopelessness	33.33	37.25	25.49	19.61
Loss of Interest/ Pleasure	54.9	60.78	50.98	13.73
Loss of Interest	50.98	56.86	45.1	17.65
Loss of Pleasure	50.98	50.98	43.14	15.69
Change in Weight	47.06	43.14	35.29	19.61
Appetite Gain	13.73	18	9.8	11.76
Appetite Loss	29.41	24	19.61	11.76
Weight Gain	11.76	14	5.88	13.73
Weight Loss	15.69	4	1.96	13.73
Sleep Disturbance	54.9	56.86	49.02	13.73
Early Insomnia	33.33	25.49	21.57	15.69
Middle Insomnia	29.41	23.53	21.57	9.8
Late Insomnia	17.65	17.65	9.8	15.69
Hypersomnia	19.61	25.49	17.65	9.8
Psychomotor Disturbance	19.61	33.33	19.61	13.73
Psychomotor Agitation	5.88	7.84	1.96	9.8
Psychomotor Retardation	15.69	27.45	15.69	11.76
Fatigue	52.94	56.86	47.06	15.69
Worthlessness/ Guilt	39.22	47.06	31.37	23.53
Worthlessness	23.53	39.22	21.57	19.61
Guilt	27.45	19.61	11.76	23.53
Concentration Difficulty	41.18	39.22	31.37	17.65
Indecisiveness	21.57	17.65	11.76	15.69
Ability to Concentrate	39.22	35.29	29.41	15.69
Suicidal Thoughts/ Behaviors	29.41	19.61	15.69	17.65
Thoughts of Death	29.41	19.61	15.69	17.65
Suicidal Ideation	23.53	13.73	11.76	13.73
Specific Plan	5.88	3.92	3.92	1.96
Past Attempt	1.96	4	1.96	1.96

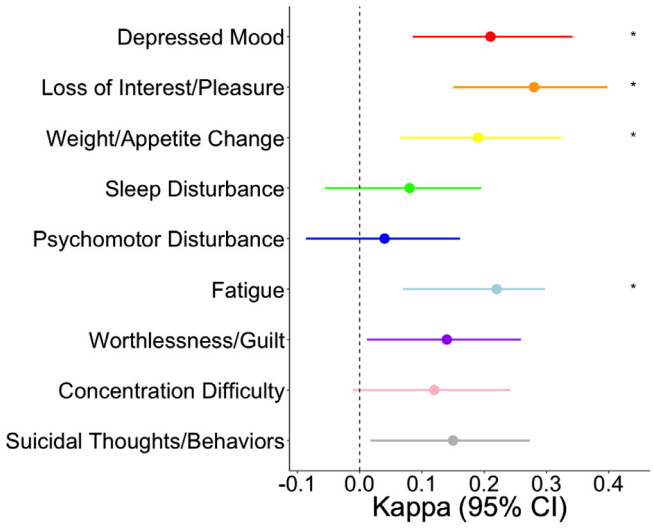

Fig. 2. Cohen's Kappa for the familial concordance of aggregated MDD symptoms.

symptoms such as the SCID. The present study found that nearly all symptoms, both aggregated and unaggregated, had significant test-retest reliability – with kappas falling in the moderate to substantial range. While aggregated and unaggregated symptoms had similar familial concordance in the slight to fair range, aggregated symptoms had slightly better estimates of familial agreement. Importantly, results also indicated that the effects were generally not influenced by differences in sex or age.

4.1. *Test-retest reliability*

Due to the symptomatic heterogeneity of MDD, questions have been raised regarding whether the different sub-parts of each symptom (e.g., psychomotor agitation vs psychomotor retardation; Walther et al., 2019) have different psychometric properties (Mazure et al. 1986). Therefore, it is important to examine the test-retest reliability of aggregated and unaggregated MDD symptoms. Our results show that the test-retest reliability of individual MDD symptoms fell into the moderate to substantial range for both aggregated and unaggregated

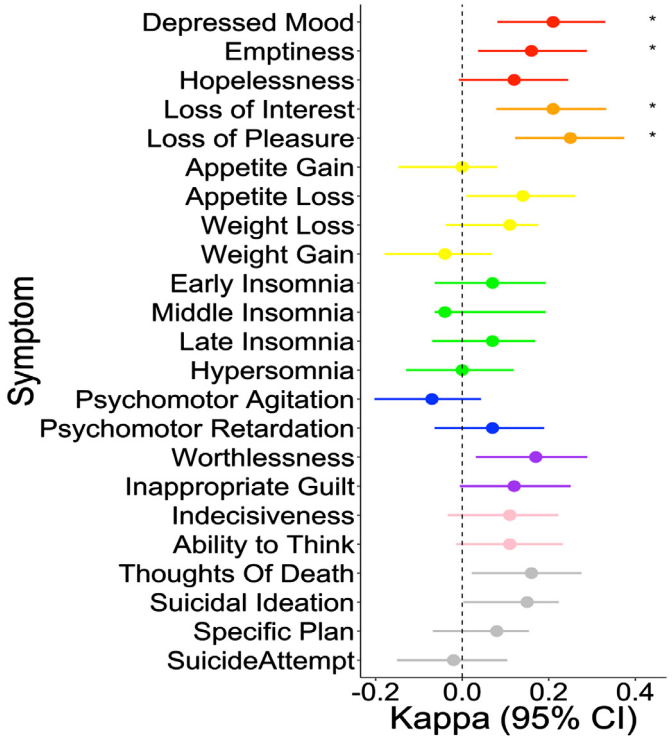

Fig. 3. Cohen's Kappa for the familial concordance of unaggregated MDD symptoms.

depressive symptoms. The results for the unaggregated symptom "specific plan for suicide" should be interpreted with caution as the sample had a low endorsed rate for this symptom. Future research should examine its test-retest reliability in samples with higher base rates of symptoms regarding suicidal behaviors, such as those with chronic or treatment resistant depression (Klein et al., 2008).

With the exception of worthlessness/guilt and weight loss/gain, the reliability of unaggregated symptoms was comparable to the corresponding aggregated symptoms, falling into the moderate to substantial range. Therefore, studies examining individual symptoms should consider the lower reliability of worthlessness/guilt and weight loss/gain when examining these symptoms separately relative to when they are aggregated. Importantly, these two aggregated symptoms may have had slightly better reliability not because their subcomponents are interchangeable, but simply because they aggregated multiple items (Sijtsma, 2009). Additionally, the poorer reliability of these (and other unaggregated) symptoms may have resulted from unclear definitions or idiosyncratic interpretations of symptoms. For example, participants/interviewers may have varied in their perception of what constitutes 'inappropriate' or 'excessive' guilt (Tilghman-Osborne et al., 2013 – see Zahn et al, 2015 for comparable discussion for worthlessness).

Our findings address questions about the reliability of single-item assessments that are increasingly being used in studies of internalizing disorders (Mitchell et al., 2007; Turon et al., 2019; Zimmerman et al., 2006). Since our results show that nearly all symptoms had fair to substantial test-retest reliability, we suggest that using items from the SCID is one way to assess individual symptoms reliably. Importantly, skip-outs were suspended, allowing for the assessment of each symptom regardless of the presence of the cardinal symptoms of MDD (or the full diagnosis of MDD). This is crucial when modeling the relationships between individual symptoms, as abiding by the skip-out rules can bias the symptom covariance estimates upon which both network models (Hoffman et al., 2019) and factor models (Kotov et al., 2018) are based.

Although these findings suggest that the SCID can assess individual MDD symptoms reliably, this study does not address other issues related to the measurement of individual symptoms. For example, the study only examined MDD symptoms as outlined in the DSM-5. While the DSM-5 criteria do a reasonable job of reflecting the prominent symptoms of depression, the list may only reflect a subset of depression-related signs and symptoms (Kendler et al., 2018; Fried et al, 2015). Indeed, non-MDD symptoms such as irritability and anger are linked to more severe and chronic depression (Judd et al., 2013). Thus, DSM-5 criteria remain an imperfect approximation of the broader syndrome. Kendler (2016) argues that diagnostic criteria are designed to index rather than describe the complete syndrome and aggregating disparate symptoms (as explored in this study) likely further impacts the reliability and validity of diagnosis. Thus, clinicians should make an effort to explore the diversity of the depressive symptoms that patients experience which includes aggregated, unaggregated, and non-DSM-5 items. Future research should also examine the reliability of individual symptoms of other, commonly comorbid disorders such as Generalized Anxiety Disorder, to facilitate transdiagnostic work on individual symptoms.

4.2. Familial concordance

In addition to reliability, our results provide insight into the validity of the aggregated and unaggregated symptoms of MDD by examining the familial concordance of MDD symptoms. A well-replicated finding in the literature is that most (if not all) psychiatric disorders are moderately heritable (Zuk et al., 2012; Sullivan et al., 2000). Although a previous study estimated the familial concordance of factors of depression symptoms (identified via exploratory factor analysis, Korszun et al., 2004), to our knowledge this study is the first to examine the familial concordance of individual MDD symptoms. Results indicated that individual symptoms of MDD had low to fair concordance between siblings, suggesting that symptoms are majorly influenced by unique (that is, unshared with their biological sibling) environmental experiences. Aggregated and unaggregated symptoms were roughly equally familial. Additionally, the significance of each symptom's familial concordance did not change when the interaction of each sibling's sex and age was included in the model, suggesting that there likely is not a sex-specific factor (e.g., genetic, hormonal, social, environmental) that strongly impacts familial concordance.

While MDD diagnosis and numerous MDD symptoms were shown to be familial, the degree to which these effects were due to genetic versus environmental factors is uncertain. Twin studies, which can quantify the relative importance of genetic, shared environmental, and unique environmental factors on MDD, offer insight into this important question. Twin studies have shown that while genetics are an important factor for certain MDD symptoms (e.g., loss of appetite, loss of libido/sexual pleasure, feelings of guilt, and hopelessness; Jang et al., 2004), most of the variance is attributable to unique environmental factors (Kendler et al., 2013) and some symptoms associated with MDD may not be heritable at all (Jang et al., 2004). Several studies have sought to identify specific genetic or environmental factors that contribute to one or more individual depressive symptoms (e.g., Keller et al., 2007; Myung et al., 2012), but further research is needed to gain insight into the etiology of depressive symptoms.

4.3. Limitations

These findings should be interpreted in the light of several limitations. First, the bootstrapped CIs for the Kappa estimates are fairly large, possibly due to the small sample size and low prevalence rates for many of the symptoms, especially the unaggregated symptoms. Replication of these results is needed. However, a power analysis revealed that the N of 51 had 95% power to detect test-retest reliabilities for moderately endorsed (> 30%) symptoms. Second, the differential variability in depressive symptoms is a potential source of biased estimates, because heavily skewed symptoms that were infrequently

endorsed are less likely to demonstrate significant statistical relationships (e.g., suicidal behaviors or psychomotor agitation in our sample). Third, the degree of accuracy for recalling symptoms of lifetime psychiatric disorders, including MDD, has been questioned (Olino et al., 2012; Takayanagi et al., 2014). According to Wells and Horwood, only 44% of participants with a lifetime diagnosis of MDD recalled a cardinal symptom of MDD, which, although concerning, also speaks to the importance of examining all 32 aggregated and unaggregated symptoms when studying lifetime MDD. In the same vein, participants may have had idiosyncratic understandings of the unaggregated symptoms. For example, participants understanding of the symptom "suicide plan" may or may not have reflected an intent to die (Linehan, 1986). Fourth, the present study only looked at one component of validity (familial concordance), and future studies should examine other aspects of validity (e.g., treatment response, functional impairment, laboratory tasks). Fifth, analyses of familial concordance examined each symptom independently, but it is possible that some symptoms share a common genetic or environmental cause. It is also possible that some symptoms are causally related (e.g., insomnia may cause fatigue; Borsboom and Cramer, 2013), which would bias the estimated familial concordance of a given symptom to the extent that it is caused by other symptoms. Sixth, test-retest reliability was not examined for the same exact participant interview, rather, it was the reliability of the same MDD episode using a different interviewer. Test retest reliability was therefore confounded with inter rater reliability; nevertheless, this is a conservative estimate of the true reliability and validity of the SCID's ability to assess individual symptoms

4.4. Conclusions

Psychopathology research has increasingly focused on individual symptoms of depression and other disorders. The majority of this work has utilized item-level analyses of assessments that were originally designed to measure latent disorders, and the psychometric properties of these single-item symptom measures are unclear. The present study found that individual MDD symptoms (as assessed by the SCID, a widely used diagnostic interview) demonstrated significant test-retest reliability in the fair to substantial range; and slight to fair familial concordance. Test-retest reliability and familial concordance was generally similar for unaggregated symptoms and their corresponding aggregated symptom. Taken together, these findings provide preliminary evidence that interview-obtained measures of individual MDD symptoms (either aggregated or unaggregated) generally have adequate test-retest reliability to assess certain lifetime depressive symptoms, but might not be able to assess all symptoms. However, further research examining the reliability and validity of both interview- and questionnaire-based measures of individual symptoms is needed.

Author statement

Ms. Kaiser was the primary author of the study and conducted the analyses along with Mr. Funkhouser. Drs. Walther and Mittal contributed to the interpretation of the data and constructs studied. Dr. Shankman was PI of the project from which the data was collected and conceptualized the study. All authors contributed to the writing of the manuscript. We wish to acknowledge Dr. Robin Mermelstein for her assistance with conceptualization and her review on a previous version of this paper.

Declaration of Competing Interest

None.

Acknowledgment

This work was supported by grants from the National Institute of Mental Health (R01MH098093 [PI: Shankman] and R01MH118741 [PI: Shankman, Mittal, and Walther]). Other support for this work was provided by the University of Illinois at Chicago (UIC) Center for Clinical and Translational Science (CCTS) award number UL1RR029879 from the National Center for Research Resources.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2020.113313.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA.
Benca, R.M., Peterson, M.J., 2008. Insomnia and depression. Sleep Med. https://doi.org/10.1016/S1389-9457(08)70010-8.
Berman, A.L., 2009. Depression and suicide. In: Gotlib, Ian H, Hammen, C.L. (Eds.), Handbook of Depression, 2nd ed. Guilford Press, New York, NY, pp. 510–530.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U., 2006. Complex networks: structure and dynamics. Phys. Rep. https://doi.org/10.1016/j.physrep.2005.10.009.
Borsboom, D., Cramer, A.O.J., 2013. Network analysis: an integrative approach to the structure of psychopathology. Annu. Rev. Clin. Psychol. https://doi.org/10.1146/annurev-clinpsy-050212-185608.
Bromet, E.J., Dunn, L.O., Connell, M.M., Amanda Dew, M., Schulberg, H.C., 1986. Long-term reliability of diagnosing lifetime major depression in a community sample. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.1986.01800050033004.
Canty, A., Ripley, B., 2013. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-9. http://Cran.R-Project.Org/Doc/Packages/.
Caspi, A., Houts, R.M., Belsky, D.W., Goldman-Mellor, S.J., Harrington, H., Israel, S., Meier, M.H., Ramrakha, S., Shalev, I., Poulton, R., Moffitt, T.E., 2014. The p factor: One general psychopathology factor in the structure of psychiatric disorders? Clin. Psychol. Sci. https://doi.org/10.1177/2167702613497473.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. https://doi.org/10.1177/001316446002000104.
Crawford, J., Cayley, C., Lovibond, P.F., Wilson, P.H., Hartley, C., 2011. Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). Aust. Psychol. https://doi.org/10.1111/j.1742-9544.2010.00003.x.
Dew, M.A., Reynolds, C.F., Houck, P.R., Hall, M., Buysse, D.J., Frank, E., Kupfer, D.J., 1997. Temporal profiles of the course of depression during treatment: predictors of pathways toward recovery in the elderly. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.1997.01830230050007.
First, M.B., Williams, J.B.W., Karg, R.S., Spitzer, R.L., 2015. Structured Clinical Interview for DSM-5 Research Version. Am. Psychiatr. Assoc., Washingt. D.C.
Fried, E.I., Nesse, R.M., 2014. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One. https://doi.org/10.1371/journal.pone.0090311.
Fried, E.I., Nesse, R.M., 2015. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. J. Affect. Disord. https://doi.org/10.1016/j.jad.2014.10.010.
Fried, E.I., Nesse, R.M., Zivin, K., Guille, C., Sen, S., 2014. Depression is more than the sum score of its parts: Individual DSM symptoms have different risk factors. Psychol. Med. https://doi.org/10.1017/S0033291713002900.
Fries, J.F., Krishnan, E., Rose, M., Lingala, B., Bruce, B., 2011. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res. Ther. https://doi.org/10.1186/ar3461.
Funkhouser, C.J., Chacko, A.A., Correa, K.A., Kaiser, A.J.E., Shankman, S.A., 2020. Unique longitudinal relationships between symptoms of psychopathology in youth: a cross-lagged panel network analysis in the ABCD study. J. Child Psychol. Psychiatry Allied Discip. https://doi.org/10.1111/jcpp.13256.
Gamer, M., Lemon, J., Fellows, I., Singh, P., 2012. Various Coefficients of Interrater Reliability and Agreement. http://Cran.R-Project.Org/Web/Packages/Irr/Irr.Pdf.
Gorka, S.M., Hee, D., Lieberman, L., Mittal, V.A., Phan, K.L., Shankman, S.A., 2016. Reactivity to uncertain threat as a familial vulnerability factor for alcohol use disorder. Psychol. Med. https://doi.org/10.1017/S0033291716002415.
Hoffman, M., Steinley, D., Trull, T.J., Sher, K.J., 2019. Estimating transdiagnostic symptom networks: the problem of ``skip outs" in diagnostic interviews. Psychol. Assess. https://doi.org/10.1037/pas0000644.
Jang, K.L., Livesley, W.J., Taylor, S., Stein, M.B., Moon, E.C., 2004. Heritability of individual depressive symptoms. J. Affect. Disord. https://doi.org/10.1016/S0165-0327(03)00108-3.
Judd, L.L., Schettler, P.J., Coryell, W., Akiskal, H.S., Fiedorowicz, J.G., 2013. Overt irritability/anger in unipolar major depressive episodes: past and current characteristics and implications for long-term course. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2013.1957.
Keller, M.C., Neale, M.C., Kendler, K.S., 2007. Association of different adverse life events with distinct patterns of depressive symptoms. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.2007.06091564.
Kendler, K.S., Aggen, S.H., 2017. Symptoms of major depression: their stability,

familiality, and prediction by genetic, temperamental, and childhood environmental risk factors. Depress. Anxiety. https://doi.org/10.1002/da.22591.
Kendler, K.S., Aggen, S.H., Flint, J., Borsboom, D., Fried, E.I., 2018. The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. J. Affect. Disord. https://doi.org/10.1016/j.jad.2017.11.032.
Kendler, K.S., Aggen, S.H., Neale, M.C., 2013. Evidence for multiple genetic factors underlying DSM-IV criteria for major depression. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2013.751.
Kendler, K.S., 2016. The phenomenology of major depression and the representativeness and nature of DSM criteria. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.2016.15121509.
Kessler, R.C., McGonagle, K.A., Swartz, M., Blazer, D.G., Nelson, C.B., 1993. Sex and depression in the national comorbidity survey i: lifetime prevalence, chronicity and recurrence. J. Affect. Disord. https://doi.org/10.1016/0165-0327(93)90026-G.
Kessler, R.C., Wai, T.C., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.62.6.617.
Khan, A.A., Gardner, C.O., Prescott, C.A., Kendler, K.S., 2002. Gender differences in the symptoms of major depression in opposite-sex dizygotic twin pairs. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.159.8.1427.
Klein, D.N., Lewinsohn, P.M., Rohde, P., Seeley, J.R., Shankman, S.A., 2003. Family study of co-morbidity between major depressive disorder and anxiety disorders. Psychol. Med. https://doi.org/10.1017/S0033291703007487.
Klein, D.N., Shankman, S.A., Rose, S., 2008. Dysthymic disorder and double depression: prediction of 10-year course trajectories and outcomes. J Psychiat. Res. 42, 408–415. https://doi.org/10.1016/j.jpsychires.2007.01.009.
Korszun, A., Moskvina, V., Brewster, S., Craddock, N., Ferrero, F., Gill, M., Jones, I.R., Jones, L.A., Maier, W., Mors, O., Owen, M.J., Preisig, M., Reich, T., Rietschel, M., Farmer, A., McGuffin, P., 2004. Familiality of symptom dimensions in depression. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.61.5.468.
Kotov, R., Waszczuk, M.A., Krueger, R.F., Forbes, M.K., Watson, D., Clark, L.A., Achenbach, T.M., Althoff, R.R., Ivanova, M.Y., Michael Bagby, R., Brown, T.A., Carpenter, W.T., Caspi, A., Moffitt, T.E., Eaton, N.R., Forbush, K.T., Goldberg, D., Hasin, D., Hyman, S.E., Lynam, D.R., Samuel, D.B., South, S.C., Markon, K., Miller, J.D., Morey, L.C., Mullins-Sweatt, S.N., Ormel, J., Patrick, C.J., Regier, D.A., Rescorla, L., Ruggero, C.J., Sellbom, M., Simms, L.J., Skodol, A.E., Slade, T., Tackett, J.L., Waldman, I.D., Widiger, T.A., Wright, A.G.C., Zimmerman, M., 2017. The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. J. Abnorm. Psychol. https://doi.org/10.1037/abn0000258.
Kovacs, M., Devlin, B., Pollock, M., Richards, C., Mukerji, P., 1997. A controlled family history study of childhood-onset depressive disorder. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.1997.01830190033004.
Krueger, R.F., 1999. The structure of common mental disorders. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.56.10.921.
Lewinsohn, P.M., Shankman, S.A., Gau, J.M., Klein, D.N., 2004. The prevalence and comorbidity of subthreshold psychiatric conditions. Psychol. Med. https://doi.org/10.1017/S0033291703001466.
Linehan, M., 1986. Suicidal people. One population or two? Ann. N. Y. Acad. Sci.
Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J., 2006. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. https://doi.org/10.1016/S0140-6736(06)68770-9.
Lovibond, P.F., Lovibond, S.H., 1995. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. Behav. Res. Ther. https://doi.org/10.1016/0005-7967(94)00075-U.
Lux, V., Kendler, K.S., 2010. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. Psychol. Med. https://doi.org/10.1017/S0033291709992157.
Markon, K.E., Chmielewski, M., Miller, C.J., 2011. The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. Psychol. Bull. https://doi.org/10.1037/a0023678.
Markon, K.E., 2010. Modeling psychopathology structure: A symptom-level analysis of Axis i and II disorders. Psychol. Med. https://doi.org/10.1017/S0033291709990183.
Mazure, C., Nelson, J.C., Price, L.H., 1986. Reliability and validity of the symptoms of major depressive illness. Arch. Gen. Psychiatry. https://doi.org/10.1001/archpsyc.1986.01800050053006.
McClintock, S.M., Husain, M.M., Greer, T.L., Cullum, C.M., 2010. Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis. Neuropsychology. https://doi.org/10.1037/a0017336.
McMakin, D.L., Olino, T.M., Porta, G., Dietz, L.J., Emslie, G., Clarke, G., Wagner, K.D., Asarnow, J.R., Ryan, N.D., Birmaher, B., Shamseddeen, W., Mayes, T., Kennard, B., Spirito, A., Keller, M., Lynch, F.L., Dickerson, J.F., Brent, D.A., 2012. Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatmentresistant depression. J. Am. Acad. Child Adolesc. Psychiatry. https://doi.org/10.1016/j.jaac.2012.01.011.
McNally, R.J., 2019. The network takeover reaches psychopathology. Beh. Brain Sci. 42, 25–26.
Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care?. A pooled analysis and meta- analysis of 22 studies. Br j Gen. Pract.
Moskvina, V., Farmer, A., Jones, I.R., Brewster, S., Ferrero, F., Gill, M., Jones, L.A., Maier, W., Mors, O., Owen, M.J., Perry, J., Preisig, M., Rietschel, M., McGuffin, P., Craddock, N., Korszun, A., 2008. Sex differences in symptom patterns of recurrent major depression in siblings. Depress. Anxiety. https://doi.org/10.1002/da.20372.
Myung, W., Song, J., Lim, S.W., Won, H.H., Kim, S., Lee, Y., Kang, H.S., Lee, H., Kim, J.W., Carroll, B.J., Kim, D.K., 2012. Genetic association study of individual symptoms in depression. Psychiatry Res. https://doi.org/10.1016/j.psychres.2011.12.037.
Nierenberg, A.A., Trivedi, M.H., Fava, M., Biggs, M.M., Shores-Wilson, K., Wisniewski, S.R., Balasubramani, G.K., Rush, A.J., 2007. Family history of mood disorder and characteristics of major depressive disorder: a STAR*D (sequenced treatment alternatives to relieve depression) study. J. Psychiatr. Res. https://doi.org/10.1016/j.jpsychires.2006.02.005.
Olino, T.M., Shankman, S.A., Klein, D.N., Seeley, J.R., Pettit, J.W., Farner, R.F., Lewinsohn, P.M., 2012. Lifetime rates of psychopathology in single versus multiple diagnostic assessments: comparison in a community sample of probands and siblings. J. Psychiatr. Res. 46, 1217–1222. https://doi.org/10.1016/j.jpsychires.2012.05.017.
Parker, G., Austin, M.P., Mitchell, P., Wilhelm, K., Boyce, P., Eyers, K., 1995. Sub-Typing Depression, I. Is Psychomotor Disturbance Necessary and Sufficient to the Definition of Melancholia? Psychol. Med. https://doi.org/10.1017/S0033291700035066.
R Core Team (2019), 2019. R: A language and environment for statistical computing. Accessed 1st April 2019.
Regier, D.A., Narrow, W.E., Clarke, D.E., Kraemer, H.C., Kuramoto, S.J., Kuhl, E.A., Kupfer, D.J., 2013. DSM-5 field trials in the United States and Canada, part II: test-retest reliability of selected categorical diagnoses. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.2012.12070999.
Revelle, W., 2016. Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA [WWW Document] R Packag. Publ. through CRAN.
Robins, E., Guze, S.B., 1970. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. Am. J. Psychiatry. https://doi.org/10.1176/ajp.126.7.983.
SCID-101 for DSM-IV: Training video for the Structured Clinical Interview for DSM-(SCID) [motion picture], 1998. In N. Y. S. P. I. Biometrics Research Department (Producer).
Shankman, S.A., Gorka, S.M., 2015. Psychopathology research in the RDoC era: unanswered questions and the importance of the psychophysiological unit of analysis. Int J of Psychophys. 98, 330–337. https://doi.org/10.1016/j.ijpsycho.2015.01.001.
Shankman, S.A., Klein, D.N., Lewinsohn, P.M., Seeley, J.R., Small, J.W., 2008. Family study of subthreshold psychopathology in a community sample. Psychol. Med. https://doi.org/10.1017/S0033291707001857.
Shankman, S.A., Lewinsohn, P.M., Klein, D.N., Small, J.W., Seeley, J.R., Altman, S.E., 2009. Subthreshold conditions as precursors for full syndrome disorders: a 15-year longitudinal study of multiple diagnostic classes. J. Child Psychol. Psychiatry Allied Discip. https://doi.org/10.1111/j.1469-7610.2009.02117.x.
Shankman, S.A., Funkhouser, C.J., Klein, D.N., Davila, J., Lerner, D., Hee, D., 2018. Reliability and validity of severity dimensions of psychopathology assessed using the structured clinical interview for DSM-5 (SCID). Int. J. Methods Psychiatr. Res. https://doi.org/10.1002/mpr.1590.
Sijtsma, K., 2009. On the use, the misuse, and the very limited usefullness of Cronbach's alpha. Psychometrika 74, 107–120.
Smoller, J.W., Finn, C.T., 2003. Family, Twin, and Adoption Studies of Bipolar Disorder. Am. J. Med. Genet. - Semin. Med. Genet. https://doi.org/10.1002/ajmg.c.20013.
Soehner, A.M., Kaplan, K.A., Harvey, A.G., 2014. Prevalence and clinical correlates of co-occurring insomnia and hypersomnia symptoms in depression. J. Affect. Disord. https://doi.org/10.1016/j.jad.2014.05.060.
Spitzer, R.L., Fleiss, J.L., 1974. A reanalysis of the reliability of psychiatric diagnosis. Br. J. Psychiatry. https://doi.org/10.1192/bjp.125.4.341.
Sullivan, P.F., Neale, M.C., Kendler, K.S., 2000. Genetic epidemiology of major depression: review and meta-analysis. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.157.10.1552.
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. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2013.3579.
Tilghman-Osborne, C., Cole, D.A., Felton, J.W., 2012. Inappropriate and excessive guilt: Instrument validation and developmental differences in relation to depression. J. Abnorm. Child Psychol. https://doi.org/10.1007/s10802-011-9591-6.
Turon, H., Carey, M., Boyes, A., Hobden, B., Dilworth, S., Sanson-Fisher, R., 2019. Agreement between a single-item measure of anxiety and depression and the Hospital Anxiety and Depression Scale: A cross-sectional study. PLoS One. https://doi.org/10.1371/journal.pone.0210111.
Uher, R., Perlis, R.H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M.Z., Souery, D., Bajs, M., Maier, W., Aitchison, K.J., Farmer, A., McGuffin, P., 2012. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. Psychol. Med. https://doi.org/10.1017/S0033291711001905.
Walther, S., Bernard, J.A., Mittal, V.A., Shankman, S.A., 2019. The utility of an RDoC motor domain to understand psychomotor symptoms in depression. Psychol. Med. https://doi.org/10.1017/S0033291718003033.
Zahn, R., Lythe, K.E., Gethin, J.A., Green, S., Deakin, J.F.W., Young, A.H., Moll, J., 2015. The role of self-blame and worthlessness in the psychopathology of major depressive disorder. J Affect. Disorder 186, 337–341. https://doi.org/10.1016/j.jad.2015.08.001.
Zimmerman, M., Ruggero, C.J., Chelminski, I., Young, D., Posternak, M.A., Friedman, M., Boerescu, D., Attiullah, N., 2006. Developing brief scales for use in clinical practice: The reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. J. Clin. Psychiatry. https://doi.org/10.4088/JCP.v67n1007.
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. https://doi.org/10.1016/j.comppsych.2014.09.007.
Zuk, O., Hechter, E., Sunyaev, S.R., Lander, E.S., 2012. The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1119675109.