The Test-Retest Reliability and Familial Concordance of Individual Symptoms of Major Depressive Disorder

BY

Ariela Kaiser
B.A., Washington University in St. Louis, 2016

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Erin Berenz, Chair, UIC Psychology
Stewart Shankman, Advisor, Northwestern Psychiatry
Robin Mermelstein, UIC Psychology
Vijay Mittal, Northwestern Psychology

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

MDD	Major Depressive Disorder
DSM-5	Diagnostic and Statistical Manual of Mental Disorders- Fifth Edition
HDRS	Hamilton Rating Scale for Depression
SSRI	Selective Serotonin Reuptake Inhibitors
RDOC	A Research Domain Criteria
SCID	Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders 5
PTSD	Posttraumatic Stress Disorder
GAD	Generalized Anxiety Disorder

SUMMARY

Major Depressive Disorder (MDD) is a heterogenous syndrome, which likely contributes to the diagnosis's questionable test-test reliability (Regier et al., 2013). MDD consists of nine core symptoms, most of which are unaggregated (e.g., sleep disturbance includes hypersomnia *or* insomnia). Individual depression symptoms may therefore exhibit different psychometric properties. This study examined two psychometric properties of the individual symptoms of MDD: (1) test-retest reliability and (2) familialness (i.e., whether they run in families). These two psychometric properties were examined for (1) the categorical diagnosis of MDD, (2) aggregated symptoms, and (3) unaggregated symptoms. Lifetime depression symptoms were measured in 504 young adults (237 sibling pairs) using an adapted version of the Structured Clinical Interview for DSM-5 (SCID; Shankman et al., 2018). Fifty-one people completed a second SCID within three weeks of their first SCID (M = 8.5 days, SD = 4.31). The test-retest reliability and familialness of each lifetime MDD symptom was evaluated using Cohen's Kappa and established conventions for agreement (Cohen, 1960). The lifetime diagnosis of MDD had substantial test-retest reliability (k = .68). The test-retest reliabilities for aggregated and unaggregated symptoms fell into the moderate to substantial range (*k's* ranged .52-.72). The lifetime diagnosis of MDD had fair familial concordance (k = .21). The familial concordance for the aggregated symptoms were highest for anhedonia (k = .28) and depressed mood (k = .21). At the unaggregated symptom level, the symptoms that comprise anhedonia—loss of interest (k = .21) and loss of pleasure (k = .25)—had the highest familial concordance and were in the fair agreement range. Given the increasing focus on the differential validity of individual MDD symptoms, our findings help illuminate whether interview-based assessments of individual symptoms reach an adequate level of reliability and validity.

I. INTRODUCTION

Major Depressive Disorder (MDD) is one the most common psychiatric disorders (Lopez, Mathers, Ezzati, Jamison, & Murray, 2006; Fried, 2014). It is a highly recurrent disorder, the leading cause of disability worldwide, a significant predictor of suicide, and often leads to severe impairment in functioning (Berman, 2009; Fried, 2014). MDD 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), an individual is diagnosed with a major depressive episode if they meet five of the nine symptoms: (1) depressed mood, (2) markedly diminished interest or pleasure, (3) increase or decrease in either weight or appetite, (4) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6) fatigue or loss of energy, (7) feelings of worthlessness or inappropriate guilt, (8) diminished ability to think or concentrate, or indecisiveness, and (9) recurrent thoughts of death or recurrent suicidal ideation. One of the symptoms must be either (1) or (2). An underlying assumption of syndromes with polythetic criteria is that the symptoms are interchangeable indicators of a disease. Major Depression is diagnosed by adding up symptoms without attention to which symptoms are present, outside of the cardinal symptoms of depressed mood or loss of interest or pleasure (although those two symptoms are viewed as interchangeable as well). In other words, the number of symptoms is emphasized rather than the nature of the symptoms.

The polythetic nature of MDD likely contributes to the poor reliability of the diagnosis. 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). Moreover, the inter-rater reliability during the field trials was worse for MDD than for the

most other disorders (e.g., post-traumatic stress disorder (PTSD) inter-rater reliability was 0.67 (CI 0.59–0.75)). The poor reliability also likely contributes to the field's difficulty identifying specific etiological factors for MDD. That is, if a construct or disorder has poor reliability, then it will be extremely difficult to identify consistent etiological factors for it (Smoller, 2003).

One explanation for the poor reliability of the MDD diagnosis is the heterogeneity of depression. In this context, heterogeneity refers to multiple individuals suffering from an extensive and varied list of psychiatric symptoms, but all meet diagnostic criteria for MDD (Olbert et al., 2014; Zimmerman et al., 2014). A study identified 1,030 unique depression symptom profiles in 3,703 individuals diagnosed with MDD (Fried & Nesse, 2015). Additionally, symptoms across episodes within an individual are often inconsistent (Oquendo et al., 2004). 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., diminished interest OR pleasure). Moreover, many of the symptoms consist of contrasting features (e.g., increase OR decrease in weight/appetite). The aggregated symptoms lead to the misconception that the compounded symptoms are binary symptoms (i.e., they do not co-occur). However, 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). Other studies have found that hypersomnia and insomnia can also co-occur in patients, as approximately one-third of patients with MDD reported both insomnia and hypersomnia (Soehner et al., 2014).

In addition to the categorical diagnosis having 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, showing this to be a valid indicator of dysfunction. The lack of reliable and valid measures for assessing individual symptoms has likely interfered with understanding the etiology and treatment progress of MDD. Additionally, frequently used scales whose objective is to assess a latent disorder do so by measuring a wide variety different depression symptoms (Shafer, 2006). For example, the gold standard measurement for depression assessment, the Hamilton Rating Scale for Depression (HDRS), does not demonstrate sufficient psychometric properties (Fried, 2014; Bagby, Ryder, Schuller, & Marshall, 2004). Moreover, many HDRS items fall into the poor inter-rater and test-retest reliability ranges and has poor content validity. Thus, in order to eventually develop psychometrically sound assessments tools for MDD, we must first fully understand the reliability and validity of individual symptoms.

A. 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 acute stressful life events such were linked to specific symptoms. For example, deaths of loved ones and romantic breakups were associated with symptoms of sadness, anhedonia, appetite loss, and increased feelings of guilt. In contrast, chronic stress had greater associations with symptoms of fatigue

and hypersomnia. Those who reported that no adverse life events precipitated their episodes reported more symptoms of fatigue, appetite gain, and thoughts of self-harm (Keller et al.,2007). Kendler and Aggen (2017) found that concordance of individual MDD symptoms among monozygotic twins was only modest and that there were separate environmental factors for different symptoms. Another study found that MDD symptoms were differentially exacerbated by chronic stress, which predicted worsening of some aggregated and unaggregated MDD symptoms (Fried et al., 2014).

MDD symptoms also differ in their impact on impairment of functioning. The symptoms that explained most of the variance included low mood, difficulty concentrating, fatigue, and loss of interest, whereas change in weight, middle insomnia, and hypersomnia were not identified as unique contributors to functional impairment (Fried & Nesse, 2014). Additionally, mid-nocturnal insomnia and hypersomnia have been shown to have less impact on functional impairment than other symptoms (Fried & Nesse, 2014). These findings support the idea that symptom sum-scores disregard qualitative differences between particular depressive symptoms.

In addition to research showing differential impact of certain MDD symptoms on impairment and etiological factors, several studies have shown that different depressive symptoms predict different treatment responses. For example, studies have shown that the presence of individual symptoms, such as sleep disturbances and hopelessness, predict a lower response to selective serotonin reuptake inhibitors (SSRI) treatment and reduced treatment efficacy (Peterson & Benca, 2008; Dew et al., 1997). Additionally, the symptom of anhedonia was associated with a poorer treatment response to SSRIs (McMakin et al., 2012). Additionally, loss of interest, diminished activity, and difficulty making decisions predicted poorer antidepressant responsivity (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 & Cramer, 2013; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). Psychopathology network analysis has become extremely popular in psychopathology research (McNally, 2019) and aims to estimate the causal relationships between individual symptoms hypothesized by the network theory (e.g., Contreras et al., 2019). 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.

B. Study Aims

Given the increased focus on individual symptoms, it is necessary to estimate MDD symptoms' psychometric properties. Thus, the study examines two psychometric properties of the symptoms of MDD: their (1) test-retest reliability and (2) familialness (i.e., the extent to which they run in families). Since Robins and Guze's (Robins & Guze, 1970) classic paper on psychiatric validity, familialness has long been considered an important criterion to determine the validity for psychiatric disorders (and, by extension, individual symptoms) (Robins & Guze, 1970; Sullivan et al., 2000). Family history of a disorder is a robust predictor of the disorder being present in probands (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 depression symptom dimensions between sibling pairs and found the highest correlations between siblings in the areas of restlessness (0.307), anxiety symptoms (0.260–0.306), loss of libido (0.295), and irritability (0.258) (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 of Korszun et al. because individuals with subthreshold MDD are a clinically significant group (Shankman et al., 2007; 2009) and only focusing on those with severe and recurrent MDD may limit the generalizability. Given the heterogeneous validity of individual MDD symptoms, certain symptoms may run in families more than others, further supporting the varied validity of individual MDD symptoms.

The present study examined these two psychometric properties (test-retest reliability and familial concordance) for (1) the categorical diagnosis of MDD, (2) aggregated symptoms (e.g., sleep disturbance overall), and (3) unaggregated symptoms (e.g., hypersomnia and insomnia).

II. METHODS

A. Participants

504 participants were recruited as part of a NIMH-funded family study (see Gorka et al., 2016; Katz, Hee, Hooker, & Shankman, 2017 for 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 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 (DASS; Lovibond & 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, Cayley, Lovibond, Wilson, & Hartley, 2011).

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., 2011; Krueger et al., 1998, 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).

Table 1
Participant Demographics and characteristics.

	Sibling Pairs (*N* = 237) Mean (SD)	Test-Retest Reliability (N=51) Mean (SD)
Age	22.3 (3.2)	22.3 (3.3)
Percent Female	64%	60%
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%

B. Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders 5 (SCID)

The SCID (First, Williams, Karg, & Spitzer, 2015) is a semi-structured clinical interview used to assess whether an individual meets criteria for any diagnoses as defined by the fifth edition of the DSM. The SCID for the present study was identical to the SCID-5, with the following changes largely made to be able to assess individual symptoms (Shankman et al., 2018). First, the instrument used a slightly different structure than the SCID-5. 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, we modified some of the skip-out rules in the SCID. Specifically, interviewers ignored all but the first "skip out" for all disorders except MDD and Generalized Anxiety Disorder (GAD), for which all the symptoms were assessed (Shankman et al., 2018). Lastly, for a number of disorder modules the SCID-5 sometimes assesses lifetime

diagnostic criteria before current criteria. For other disorder modules, the order is reversed. The following modules were administered in the current study: MDD, Alcohol Use Disorder, Substance Use Disorder, PTSD, Panic Disorder, Agoraphobia, Social Anxiety Disorder, Specific Phobia, Obsessive Compulsive Disorder, GAD, Anorexia, Bulimia, Binge Eating Disorder, and the bipolar and psychotic screening modules. Doctoral students and bachelor's level research assistants were trained to criterion on the SCID and were supervised by a licensed clinical psychologist. Inter-rater agreement was in the fair to substantial ranges for lifetime diagnoses (*k's* =.46–.87) (Shankman et al., 2018; Shrout, 1998). 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.

Interviewers were trained to criterion by viewing the SCID-101 training videos (SCID-101, 1998), overserving two or three SCID interviews with an experienced interviewer, and completing three SCID interviews (observed by an advanced interviewer) in which diagnoses in were in full agreement with those of the observer.

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).

C. 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 & Ripley,

2017). 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. In order to demonstrate that our randomization of the sibling pair orders was not biased, we also performed logistic regressions (i.e., sibling 1 predicting sibling 2) to show our familial associations are robust and to ensure that the results were not biased by the randomization. We determined that a symptom effect was statistically significant only if both the Kappa and the regression were significant for a particular symptom/disorder.

Sex and age have been shown to impact the clinical presentation of MDD. Women are approximately two times more likely than 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, 1996; Lux &Kendler, 2010). For example, hypersomnia is less common in younger people (Kovacs, 1996) while somatic complaints are more common in younger people (McCauley et al., 1991). 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 concordance covaried for each sibling's sex and age and the interaction of the two sibling's 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 is age- and sex-dependent, then the results would no longer remain when including these covariates.

III. RESULTS

A. Test-Retest Reliability

For the 51 participants who completed two SCID, the lifetime diagnosis of MDD had substantial test-retest reliability (k = .68). The test-retest reliabilities for aggregated symptoms fell into the moderate range and were highest for anhedonia (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) (see Figure 1). At the unaggregated symptom level, nearly all unaggregated symptoms had fair to substantial test-retest reliability (see Figure 2). 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) (see Table 2).

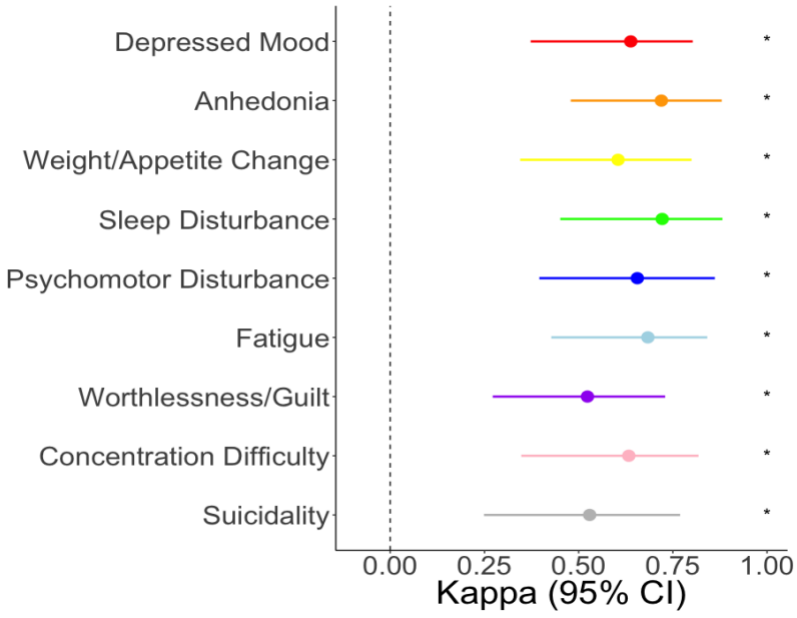

Figure 1.

Cohen's Kappas for the test-retest reliability of aggregated MDD symptoms.

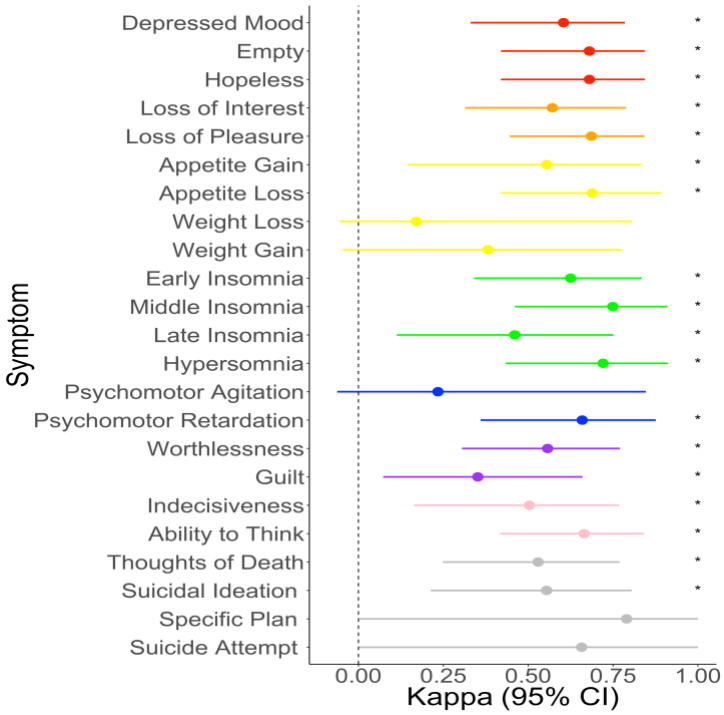

Figure 2.
Cohen's Kappas for the test-retest reliability of unaggregated MDD symptoms

Table 2
Aggregated and Unaggregated Symptom Endorsement at Time 1 and Time 2

Symptom	% Endorsed at Time 1	% Endorsed at Time 2
Depressed Mood	56	58
Low Mood	54	54
Emptiness	43	43
Hopelessness	33	37
Anhedonia	54	61
Loss of Interest	50	50
Loss of Pleasure	50	56
Change in Weight	47	43
Eat More	13	17
Eat Less	29	23

Weight Gain	11	13
Weight Loss	15	3
Sleep Disturbance	54	56
Early Insomnia	33	25
Middle Insomnia	29	23
Late Insomnia	17	17
Hypersomnia	19	25
Psychomotor Disturbance	19	33
Psychomotor Agitation	5	7
Psychomotor Retardation	15	27
Fatigue	52	56
Worthlessness	60	52
Worthlessness	23	39
Guilt	27	19
Low Self Esteem	52	49
Decision Making	41	39
Concentrate	39	35
Indecisive	21	33
Suicidal Thoughts/Behaviors	29	19
Thoughts of Death	29	19
Ideation	23	13
Specific Plan	5	3
Past Attempt	1	3

B. Familial Concordance

In the sample of 237 sibling pairs, there was fair familial concordance for the lifetime categorical MDD diagnosis (k = .21). The familial concordance for the aggregated symptoms is presented in Figure 3 and were highest for anhedonia (k = .28) and depressed mood (k = .21). At

the unaggregated symptom level, the symptoms that comprise anhedonia—loss of interest (*k* = .21) and loss of pleasure *(k* = .25)—had the highest familial concordance and were in the fair agreement range (see Table 3). The unaggregated symptoms with the lowest familial concordance included sleep disturbance (*k* = .08) and psychomotor disturbance (*k* = .04) (see Table 4). The unaggregated symptoms' familial concordances are presented in Figure 4.

Table 3
Aggregated Symptom (siblings)

	Kappa	Lower CI	Upper CI	% Endorsed
Depressed Mood	.21	.08	.33	421%
Emptiness	.19	.05	.32	31%
Hopelessness	.12	-.01	.26	32%
Loss of Interest	.19	.05	.31	36%
Loss of Pleasure	.24	.1	.37	34%
Appetite Gain	.11	.01	.24	9%
Appetite Loss	.02	-.06	.14	14%
Weight Gain	.1	-.02	.25	8%
Weight Loss	.03	-.07	.22	21%
Early Insomnia	-.02	-.07	.11	12%
Middle Insomnia	.04	-.07	.19	9%
Late Insomnia	-.01	-.09	.2	19%
Hypersomnia	-.02	-.1	.11	8%
Psychomotor Agitation	.02	-.07	.02	10%
Psychomotor Retardation	-.07	-.01	-.05	18%
Worthlessness	.04	-.07	.2	22%
Guilt	.18	-.01	.3	27%
Indecisiveness	.15	.03	.29	15%
Ability to concentrate	.01	-.09	.14	32%
Thoughts of Death	.13	.01	.3	20%
Thoughts of Suicide	.1	-.03	.31	16%
Specific Plan	.13	.01	.34	6%
Past Attempt	.15	-.03	.41	3%

Table 4
Unaggregated Symptom (siblings)

	Kappa	Lower CI	Upper CI	% Endorsed
Depressed Mood	.21	.09	.33	45%
Anhedonia	.28	.15	.4	39%
Appetite/Weight Change	.19	.08	.34	32%
Sleep Disturbance	.04	-.08	.19	16%
Psychomotor Disturbance	.08	-.05	.21	41%
Fatigue	.22	.1	.35	38%
Worthlessness/Guilt	.14	.01	.26	29%
Concentration Difficulty	.12	-.01	.24	33%
Suicidal Thoughts/Behaviors	.15	.01	.3	20%

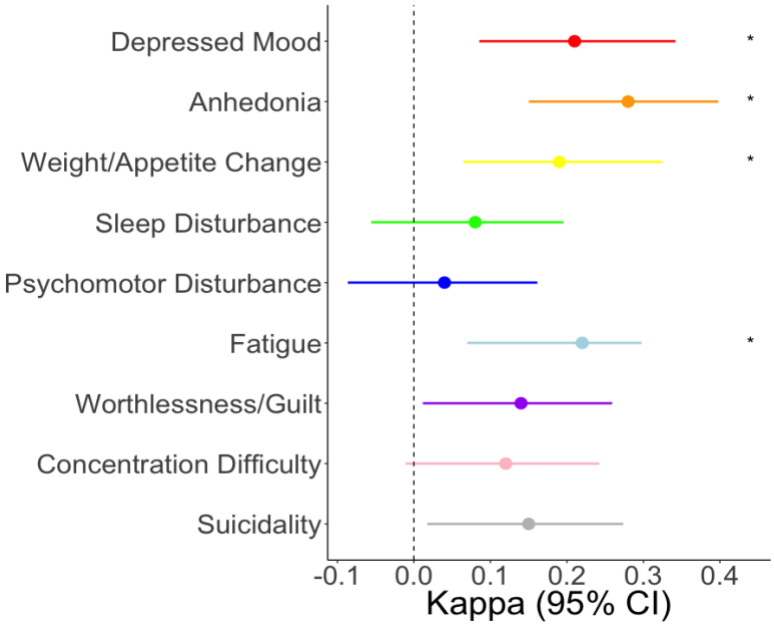

Figure 3.
Cohen's Kappa for the familial concordance of aggregated MDD symptoms.

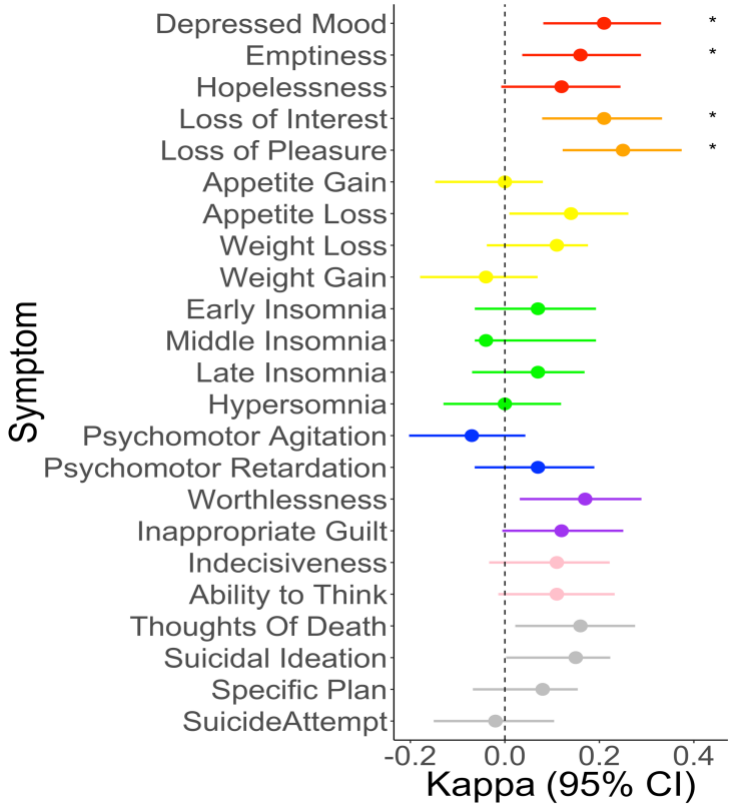

Figure 4.
Cohen's Kappa for the familial concordance of unaggregated MDD symptoms.

C. Impact of Sex and Age on Familialness

Results of logistic regressions examining familial associations 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 = .796, 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 take into account sex and age differences.

IV. DISCUSSION

Research has increasingly studied the etiology, course, and treatment responsiveness of individual MDD symptoms and not just the overall disorder. Many of these studies have assessed symptoms using single items from structured interviews. However, the psychometric properties of these assessments (e.g., test-retest reliability) are unclear. Therefore, it is necessary to study the reliability and validity of tools that can measure individual symptoms such as the SCID. The present study found that all symptoms, both aggregated and unaggregated, had fair to substantial test-retest reliability. 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 sex.

A. Test-Retest Reliability

The present study demonstrated better test-retest reliability for the categorical MDD diagnosis than the DSM-5 field trials (Regier et al., 2013*)*, which showed lower diagnostic reliability than past field trials (Spitzer et al., 1979; Brown et al., 2001), resulting in criticism about the reliability of the DSM-5 (Freedman et al., 2013). Without diagnostic reliability, it is not possible to have valid diagnoses (Nelson-Gray, 1991). The DSM- 5 field trials, however, did not use a standardized clinical interview across sites or interviewers (i.e., interviewers used their clinical interview of choice and a symptom checklist to document the presence or absence of symptoms needed to support their diagnosis) and this likely contributed to the inter-rater reliability falling into the questionable range (k = 0.28) for the MDD diagnosis in the DSM-5 field trials. To address this concern, in the present study, each interviewer administered a standardized interview (e.g., the modified SCID) in order to focus on the reliability of the MDD

DSM-5 criterion. Our results show that a lifetime diagnosis of MDD had substantial test-retest reliability (k = .68). These changes in the diagnostic interviews likely led to the significantly better test-retest reliability than the DSM-5 field trials.

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) have different properties. 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 depressive symptoms. The aggregated symptoms with the highest test-retest reliability were anhedonia (k = .71) and sleep disturbances (k = .72). The unaggregated symptoms with the highest test-retest reliability were middle insomnia (k = .75), hypersomnia (k = .72), and specific plan for suicide (k =. 78). These findings support the increased importance of examining deficits in the positive valance construct (i.e., engaging in behaviors that lead to achievement and satisfaction from rewards) and the role of circadian and sleep rhythms in the pathogenesis and treatment of depression (Germain & Kupfer, 2008; Olino, 2016). The results for the unaggregated symptom "specific plan for suicide" should be interpreted with caution as the sample had a low endorsed rate of suicidal thoughts and behaviors. Future research should examine the test-retest reliability in samples with higher base rates of symptoms regarding suicidality.

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 examined

unaggregated relative to when they are aggregated. While these aggregated symptoms had slightly better reliability compared to the corresponding unaggregated symptoms, this could be explained by the possibility that aggregated symptoms were more reliable simply because they are aggregations of multiple items.

Our findings address the criticism raised against studies testing the network theory of psychopathology that single items may have poor reliability (Forbes et al., 2017). 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 SCID only measures DSM-5 MDD symptoms despite prior research showing that current criteria for MDD potentially reflect only a subset of MDD signs and symptoms (Kendler et al., 2018; Fried et al, 2015). For example, non-MDD symptoms such as irritability and anger are linked to more severe and chronic depression (Judd, Schettler, Coryell, Akiskal, & Fiedorowicz, 2013). Future research should also examine the reliability of individual symptoms of other, commonly comorbid disorders such as GAD to facilitate transdiagnostic work on individual symptoms. Furthermore, studies examining individual symptoms of MDD use a variety of questionnaires (e.g., Beck Depression Inventory; Beck et al., 1960) that are often assumed to consist of indistinguishable indicators of the

disorder, when in reality different measures of MDD often measure different constructs entirely. Fried (2017) found that the seven most common depression scales include 52 distinct symptoms, many of which are not included in the DSM-5 criteria for MDD.

B. Validity: 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, Hechter, Sunyaev, & Lander, 2012; Sullivan, Neale, & Kendler, 2000). One study that examined if the individual symptoms of MDD are familial (Korszun et al., 2004) looked at associations of depression symptom dimensions between siblings with severe recurrent depression. Our results expanded on this by exploring aggregated and unaggregated symptoms in sibling pairs from a more representative community sample. Our analyses between sibling pairs demonstrate that individual symptoms of MDD had low to fair concordance between siblings and that both the aggregated and unaggregated levels were roughly equally familial, with the highest familial concordance being only in the fair range (k = .28). Familial concordance for MDD symptoms was only slight to fair, suggesting they are majorly influenced by unique (that is, unshared with their biological sibling) environmental experiences.

While the diagnosis of MDD and numerous MDD symptoms were shown to be familial, the degree to which these effects were due to genetic vs 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 MDD and its symptoms, most of the variance are attributable to unique environmental factors (Kendler et al., 2013), such as loss of a

loved one (Lux & Kendler, 2010). Jang et al. (2014) found the aspects of MDD that were significantly heritable were physiological functions (e.g., loss of appetite, loss of libido/sexual pleasure), feelings of guilt and hopelessness, and low positive affect. In contrast, other symptoms associated with MDD, such as negative affect, nausea and headaches, or tearfulness, do not appear to be heritable (Jang et al., 2014). The present study expanded on Jang et al. by assessing MDD symptoms through a structured clinical interview (i.e., SCID) rather than measuring symptoms through self-report questionnaires (i.e., The Beck Depression Inventory; Beck et al., 1960).

Family and twin studies have demonstrated the contribution of environmental and genetic variance broadly, but they rarely identify specific genetic factors and environmental factors that lead to the risk of MDD onset. Therefore, studies have searched to identify specific genetic variables involved in MDD risk. Analysis of candidate genes, genome-wide association analysis, genome-wide sequencing, and various other methodological approaches have been used, however, in most cases, only a small number of genes have been shown to be associated MDD development and risk. In addition to specific genetic variables, specific environmental variables, such as severe loss, has been linked to the onset of MDD (Finlay-Jones & Brown, 1981). Caspi et al. (2003), reported an interaction effect of life stress and 5-HTTLPR genotype as a risk of depression. As a result of these findings, the importance of gene by environment interactions in the etiology of depression has been widely accepted (Eaves et al., 2006). MDD symptoms vary in important aspects, such as genetic association, and are not merely interchangeable indicators of a latent disease.

It is important to consider that the present study only looked at one component of validity –familial association—however, there are multiple other methods used to examine validity.

Increasingly, research has focused on the differential validity of individual depressive symptoms to predict a variety of outcome measures. For example, the presence of sleep disturbances and hopelessness predict a lower response to SSRI treatment and reduced depression treatment efficacy (Peterson & Benca, 2008; Dew et al., 1997). Other indicators of symptom validity (e.g., correlating specific MDD symptoms with stressful life events, treatment response, or functional impairment) should be examined in future work.

C. Impact of Sex and Age Differences on Psychometric Properties of Individual MDD Symptoms

In the present study, all symptoms remained significantly correlated whether or not the interaction of each sibling's sex and age was included in the model. This suggests that there likely is not a sex-specific factor (e.g., genetic, hormonal, social, environmental) that strongly impacts familial concordance. However, as mentioned above, familial concordance is only one component of validity. While sex did not influence our results, sex differences in MDD play an important role in other studies that tested the differential validity of specific symptoms. For example, Lux & Kendler found sex was differentially associated with unique depression symptom profiles (Lux & Kendler, 2010). Future studies should examine sex as a moderator for individual symptom concordance in sibling pairs. Research on sex as moderators would be particularly valuable in identifying individuals who are at increased risk of particular MDD symptom profiles. This would have important clinical implications as these different symptom profiles could be related to differential treatment responsiveness.

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 sample size and low prevalence rates for many of the symptoms, especially the unaggregated symptoms. Second,

a potential sources of biased estimates is the differential variability in depressive symptoms. Heavily skewed symptoms that were infrequently endorsed are less likely to demonstrate a significant statistical relationships (e.g., suicidality in our sample). Third, the degree of accuracy for recalling symptoms of lifetime psychiatric disorders, including MDD, has been questioned (Takayanagi et al., 2014; Wells & Horwood, 2004). According to Wells & 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. Fourth, each symptom was examined independently; however, some symptoms may be dependent on others. Psychopathological symptom network theory views mental disorders as clusters of related symptoms due to causal links between (Cramer et al., 2010). The symptom network theory does not understand MDD as a latent disease. MDD is viewed as a disorder comprised of causal connections of symptoms to each other (Borsboom & Cramer, 2013). It is therefore recommended that this theory be considered when interpreting results examining individual symptoms of MDD.

Lastly, latent variable models deem the nine aggregated MDD symptoms as equally central to the disorder. However, the assumption that different symptom are equivalent is problematic. Perhaps a shift toward symptom network models to examine causal links between symptoms will remedy the issues of using symptom sum-scores to diagnose MDD. Sum-scores emphasize the number rather than the nature of symptoms, which obscures information about unique characteristics of individual symptoms. To gain further insight into other possible etiologic influences that impact symptoms displayed by those affected with MDD, future research efforts should examine if different symptoms in sibling pairs have different pathophysiological mechanisms. While two siblings may have experienced the same symptom,

the etiology of that symptom may differ drastically (e.g., one sibling may experience depressed mood caused by a romantic breakup, whereas a second sibling might experience depressed mood caused by a neurobiological mechanism). Understanding the various etiological factors of individual MDD symptoms such as early stressful life events (Gilmer & McKinney, 2003; Gutman & Nemeroff, 2003), specific personality traits (Angst & Clayton, 1986; Kendler, Kuhn & Prescott, 2004), number of previous episodes, age of onset (Janssen & Beekman et al., 1995; Colman et al., 2011), and family history (Nierenberg et al., 2007), would improve the development of targeted prevention and intervention programs. Further examination is required to identify which specific environmental factors are associated with individual aggregated or unaggregated symptoms.

V. CITED LITERATURE

Angst, J. (1999). Major depression in 1998: are we providing optimal therapy? The Journal of clinicalpsychiatry, 60 Suppl 6, 5–9.

Angst, J., & Clayton, P. (1986). Premorbid personality of depressive, bipolar, and schizophrenic patients with special reference to suicidal issues. *Comprehensive Psychiatry*.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). Reviews and Overviews The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? American Journal of Psyc, 161(12), 2163–2177.

Baumeister, H., & Parker, J. D. (2012). Meta-review of depressive subtyping models. Journal of affective disorders, 139(2), 126–40. doi:10.1016/j.jad.2011.07.015

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry* https://doi.org/10.1001/archpsyc.1961.01710120031004

Berman, A. L. (2009). Depression and suicide. In Ian H Gotlib & C. L. Hammen (Eds.), Handbook of depression (2nd ed., pp. 510–530). New York, NY: Guilford Press.

Benca, R. M., & Peterson, M. J. (2008). Insomnia and depression. Sleep medicine, 9 Suppl 1,S3 9.doi:10.1016/S1389-9457(08)70010-8

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. Physics Reports, 424(4-5), 175–308. doi:10.1016/j.physrep.2005.10.009

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. Psychological bulletin, 110(2), 305–314.

Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*

Brown TA, Di Nardo PA, Lehman CL, Campbell LA: Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. J Abnorm Psychol 2001; 110:49–58

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*. https://doi.org/10.1159/000497425

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*.

Colman, I., Naicker, K., Zeng, Y., Ataullahjan, A., Senthilselvan, A., & Patten, S. B. (2011). Predictors of long-term prognosis of depression. *CMAJ*. https://doi.org/10.1503/cmaj.110676

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). *Australian Psychologist*.

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. Archives of general psychiatry, 54(11), 1016 24.

Finlay-Jones, R., & Brown, G. W. (1981). Types of Stressful Life Event and the Onset of Anxiety and Depressive Disorders. *Psychological Medicine*. https://doi.org/10.1017/S0033291700041301

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-V research version*. *American Psychiatric Association*.

Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., … Yager, J. (2013). The initial field trials of DSM-5: New blooms and old thorns. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2012.12091189

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are "good" depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2015.09.005

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR∗D study. Journal of Affective Disorders, 172, 96 102.https://doi.org/10.1016/j.jad.2014.10.010

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

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000276

Germain, A., & Kupfer, D. J. (2008). Circadian rhythm disturbances in depression. Human Psychopharmacology. https://doi.org/10.1002/hup.964

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. Psychological Medicine, 46(16), 3349 3358.

Gutman, D. A., & Nemeroff, C. B. (2003). Persistent central nervous system effects of an adverse early environment: Clinical and preclinical studies. In *Physiology and Behavior*. https://doi.org/10.1016/S0031-9384(03)00166-5

Gilmer, W. S., & McKinney, W. T. (2003). Early experience and depressive disorders: Human and non-human primate studies. *Journal of Affective Disorders*. https://doi.org/10.1016/S0165-0327(03)00046-6

Hofmann, S. G., Curtiss, J., & McNally, R. J. (2016). A Complex Network Perspective on Clinical Science. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691616639283

Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of Affective Disorders*. https://doi.org/10.1016/S0165-0327(03)00108-3

Janssen, J., Beekman, A. T. F., Comijs, H. C., Deeg, D. J. H., & Heeren, T. J. (2006). Late-life depression: The differences between early- and late-onset illness in a community-based sample. *International Journal of Geriatric Psychiatry*. https://doi.org/10.1002/gps.1428

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

Katz AC, Hee D, Hooker CI, Shankman SA. A Family Study of the DSM-5 Section III Personality Pathology Model Using the Personality Inventory for the DSM-5 (PID5). *Journal of Personality Disorders*.

Kaymaz, N., van Os, J., Loonen, A. J. M., & Nolen, W. A. (2008). Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta-analysis of placebo-controlled randomized trials. The Journal of clinical psychiatry, 69(9), 1423– 36

Kendler, Kenneth S, Aggen, S. H., & Neale, M. C. (2013). Evidence for Multiple Genetic FactorsUnderlying DSM-IV Criteria for Major Depression. American Journal of Psychiatry, 70(6), 599 607. doi:10.1001/jamapsychiatry.2013.751

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The Interrelationship of Neuroticism, Sex, and Stressful Life Events in the Prediction of Episodes of Major Depression. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.161.4.631

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine*. https://doi.org/10.1017/S0033291706009524

Kendler, K. S., & Aggen, S. H. (2017). Symptoms of major depression: Their stability, familiality, and prediction by genetic, temperamental, and childhood environmental risk factors. *Depression and Anxiety*. https://doi.org/10.1002/da.22591

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of general psychiatry, 62(6), 617–27. doi:10.1001/archpsyc.62.6.617

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. *Journal of Affective Disorders*. https://doi.org/10.1016/0165-0327(93)90026-G

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. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.159.8.1427

Kotov, R., Waszczuk, M. A., Krueger, R. F., Forbes, M. K., Watson, D., Clark, L. A., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*. https://doi.org/10.1037/abn0000258

Korszun, A., Moskvina, V., Brewster, S., Craddock, N., Ferrero, F., Gill, M., … McGuffin, P. (2004). Familiality of Symptom Dimensions in Depression. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.61.5.468

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. L. (2006). Global Burden of Disease and Risk Factors. Washington (DC): World Bank.

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: A validation study of the DSM-IV symptomaticcriteria. *Psychological Medicine*. https://doi.org/10.1017/S0033291709992157

Mazure, C. M. (1998). Life Stressors as Risk Factors in Depression. Clinical Psychology: Science and Practice, 5(3), 291–313. doi:10.1111/j.1468-2850.1998.tb00151.x

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, 24(1), 9–34. doi:10.1037/a0017336

McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., … Brent, D. A. (2012). Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment resistant depression. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2012.01.011

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J., Coryell, W., … Maser, J.D. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. The American journal of psychiatry, 156(7), 1000–6.

Moskvina, V., Farmer, A., Jones, I. R., Brewster, S., Ferrero, F., Gill, M., … Korszun, A. (2008). Sex differences in symptom patterns of recurrent major depression in siblings. *Depression and Anxiety*. https://doi.org/10.1002/da.20372

Nierenberg, A. A., Trivedi, M. H., Fava, M., Biggs, M. M., Shores-Wilson, K., Wisniewski, S. R., … 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. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2006.02.005

Olbert C. M., Gala G. J., Tupler L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *J. Abnorm. Psychol.* 123 452–462 10.1037/a0036068

Olino, T. M. (2016). Future Research Directions in the Positive Valence Systems: Measurement, Development, and Implications for Youth Unipolar Depression. Journal of Clinical Child and Adolescent Psychology. https://doi.org/10.1080/15374416.2015.1118694

Oquendo, M. A., Barrera, A., Ellis, S. P., Li, S., Burke, A. K., Grunebaum, M., … Mann, J. J. (2004).Instability of symptoms in recurrent major depression: a prospective study. The American journal of psychiatry, 161(2), 255–61. doi:10.1176/appi.ajp.161.2.255

Peterson, M. J., & Benca, R. M. (2008). Sleep in Mood Disorders. Sleep Medicine Clinics, 3(2), 231 249. doi:10.1016/j.jsmc.2008.01.009

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 CategoricalDiagnoses. *The American journal of psychiatry,170(1), 59–70.*

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: its application toschizophrenia. The American Journal of Psychiatry. https://doi.org/10.1176/ajp.126.7.983

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological

phenomena. New Ideas in Psychology, 31(1), 43–53. doi:10.1016/j.newideapsych.2011.02.007

Soehner, A. M., Kaplan, K. A., & Harvey, A. G. (2014). Prevalence and clinical correlates of co-occurring insomnia and hypersomnia symptoms in depression. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2014.05.060

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). *International Journal of Methods in Psychiatric Research*. https://doi.org/10.1002/mpr.1590

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. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/j.1469-7610.2009.02117.x

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. *Psychological Medicine*. https://doi.org/10.1017/S0033291707001857

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*. https://doi.org/10.1191/096228098672090967

Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Numberger JI, Sullivan PF. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. Lancet. 2013;381:13711379. doi: 10.1016/s0140 6736(12)62129-1

Spitzer, R. L., Forman, J. B. W., & Nee, J. (1979). DSM-III field trials: I. Initial interrater diagnostic reliability. American Journal of Psychiatry.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry, 157 (10) 1552 1562*

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

Uher, R, Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., … McGuffin, P. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interestactivity symptoms. Psychological medicine, 42(5), 967–80. doi:10.1017/S0033291711001905

Wells, J. E., & Horwood, L. J. (2004). How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. *Psychological Medicine*. https://doi.org/10.1017/S0033291703001843

Zimmerman M., Ellison W., Young D., Chelminski I., Dalrymple K. (2014). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Compr.Psychiatry.* 56 29–34 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. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1119675109

VI. VITA

NAME: Ariela J. Kaiser

EDUCATION: B.A., Psychological and Brain Sciences; Women, Gender, and Sexuality Studies, Washington University in St. Louis, St. Louis, Missouri, 2016

HONORS: Postbaccalaureate Intramural Research Training Award (IRTA), National Institutes of Health, 2016

Highest Distinction, Psychological and Brain Sciences Major, 2016

Magna Cum Laude

Phi Beta Kappa Honors Society

PUBLICATIONS:

Keren, H., O'Callaghan, G., Vidal-Ribas, P., Buzzell, G. A., Brotman, M. A., Leibenluft, A., Pan, P., Meffert, L., **Kaiser, A.,** Wolke, S., Luby, J., Pine, D. S., & Stringaris, A. (2017). Reward processing in depression: A conceptual and metanalytic review across behavioral, electrophysiological and imaging studies. *Am J Psychiatry (in press).*

Vidal-Ribas, P., Brotman, M. A., Salum, G. A., **Kaiser, A**., Meffert, L., Pine, D. S., Leibenluft, E., & Stringaris, A. (2018). Deficits in emotion recognition increase the risk of depressive symptoms in youth with severe irritability. *Depression and Anxiety (in press)*

O'Callaghan, G., Wolke, S., **Kaiser, A.,** Mehta, M., Young, A. & Stringaris, A. (2017). A randomized, double- blind, placebo-controlled-crossover design investigating the impact of a pharmacological challenge on neural mechanisms of response inhibition in depression. *Biological Psychiatry (in press)*

Towbin, K., Vidal-Ribas, P., Brotman, M. A., Pickles, A., Miller,V., **Kaiser, A.,** Stringaris, A. (2019). A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability. *Journal of American Academy of Child Adolescent Psychiatry (in press)*

CONFERENCE POSTERS AND PRESENTATIONS:

Kaiser, A., Shenberger, E., Correa,K., Funkhouser, C. J., & Shankman, S.(2019, May). *Familial association between reward anticipation frontal asymmetry and depressive symptoms.* Poster presented at the Society for Research in Psychopathology Annual Convention, Buffalo, NY

Kaiser, A., Funkhouser., Shankman., S. (2019, March). *The Test-Retest Reliability and Familial Concordance of the Individual Symptoms of Major*

Depressive Disorder. Poster presented at the Anxiety and Depression Association of American Conference (ADAA) Annual Convention, Chicago, IL

Kaiser, A., O'Callaghan, G., Wolke, S., Mehta, M., Young, A., Stringaris, A. (2017, August). *Stop or go? Characterizing inhibition during a pharmacological challenge in depression.* Poster presented at the American Psychological Association (APA) Annual Convention, Washington, D.C.

Kaiser, A., Vidal-Ribas, P., Wolke, S., Stringaris, A. (2017, May*) Developing a screening measure of anhedonia in adolescents*. Poster Presentation at the NIH Postbac Poster Day, Bethesda, MD

Kaiser, A., (2016, May). *The effects of race on children's reasoning about peer-based social exclusion.* Poster presented to the Washington University in St. Louis Honors Thesis Committee, St. Louis, MO

PROFESSIONAL AFFILIATIONS: Society for Research in Psychopathology, Association for Behavioral and Cognitive Therapies, Anxiety and Depression Association of America