VALIDATING A DEFINITION OF RELAPSE IN ANOREXIA NERVOSA

Emily Caroline Walsh

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology and Neuroscience in the College of Arts and Sciences.

Chapel Hill
2021

Approved by:

Anna Bardone-Cone

Don Baucom

Mary Hill

© 2021
Emily Caroline Walsh
ALL RIGHTS RESERVED

ABSTRACT

Emily Caroline Walsh: Validating a Definition of Relapse in Anorexia Nervosa
(Under the direction of Anna Bardone-Cone)

There is currently no consensus on the definition of "relapse" in anorexia nervosa (AN). This study sought to extend recent theoretical work exploring relapse in AN by comparing eight theory-driven definitions of relapse constructs comprised of combinations of physical (i.e., weight status), cognitive (e.g., shape and weight concerns), and behavioral (e.g., bingeing and purging) symptoms, ranging from full-threshold DSM-5 AN to the presence of *only* disordered eating cognitions or *only* disordered eating behaviors in the absence of other symptoms. Among 26 individuals with a lifetime history of AN who had reached recovery, five participants met criteria for relapse over the course of the study. We discuss patterns observed in these individuals compared to those who did not relapse. A small sample size limits the conclusions that can be drawn, though the authors encourage others to continue expanding upon this work with a larger sample size, ideally derived from multi-site studies.

TABLE OF CONTENTS

LIST OF FIGURES……..…vi

LIST OF TABLES…….…..vii

INTRODUCTION……..…1

Relapse Constructs Overview…….....4

DSM-5 Diagnostic Criteria Relapse Constructs…….....4

Khalsa et al. (2017) Relapse Constructs……..…...6

Cognitive and Behavioral Relapse Constructs……...….7

Eating Disorder Constructs Potentially Related to Relapse……9

Non-Eating Disorder Constructs Potentially Related to Relapse……..10

Study Aims and Hypotheses……..14

METHODS……15

Participants and Recruitment…….15

Study Procedures……...…16

Eating Disorder Recovery Status Definition and Assessment Tools……16

Operationalization of Relapse Definition Components……18

Assessment Tools Used to Measure Relapse Definition Components…….20

Measures for Assessing Validity of Relapse Definitions…….21

DATA ANALYTIC PLAN……..28

RESULTS…….29

Description of Sample……...29

Aim 1: Comparing Rates of Relapse Across Relapse Definitions……30

Aim 2: Exploring Rates of Agreement Between Relapse Conditions……...31

Aim 3: Comparing Relapsed vs. Not Relapsed Individuals' Scores on Clinically Relevant Measures……31

DISCUSSION……34

Case Study…….....38

Strengths and Limitations……..40

Future Directions……...……43

Conclusion……45

APPENDIX 1: DSM-5 EATING DISORDERS DIAGNOSTIC CRITERIA……..47

APPENDIX 2: FIGURE 1- PROCEDURE TIMELINE……..49

APPENDIX 3: TABLES……..50

REFERENCES……63

LIST OF FIGURES

Figure 1 – Procedure Timeline……..49

LIST OF TABLES

Table 1 - Relapse Constructs, Operationalization of Symptoms, and Assessment Tools Used to Assess Symptoms……..50

Table 2 - Measures Used to Assess Validity of Relapse Definitions and Internal Consistency Data in the Study Sample…….53

Table 3 - Bardone-Cone et al. (2010) Eating Disorder Recovery Definition and Assessment Tools…....55

Table 4 - Relapse Definitions Met by Individual Participants Across Both Study Timepoints…56

Table 5 - Relapse Definitions Met at Time 2 and Time 3, Reported Separately……...57

Table 6 - Cohen's Kappa for Relapse Definitions at T2……58

Table 7 - Cohen's Kappa for Relapse Definitions at T3……59

Table 8 - Independent Samples t-tests Between Individuals Who Met Relapse Definitions and Those Who Did Not Relapse Over the Study Period……..…60

Table 9 - Contingency Tables for Measures for Which There Are Clinically Meaningful Cutoff Scores…….62

Introduction

Anorexia nervosa (AN) is a serious psychiatric disorder characterized by a protracted course of illness and high mortality rates (Arcelus et al., 2011). Although AN is notoriously resistant to psychotherapeutic treatments, recovery is possible (Steinhausen, 2002). However, many, if not most, relapse back into illness within the first year after intervention and weight restoration (Pike, 1998) and are unable to sustain their "recovered" status. It is estimated that the window of four to 16 months after successful treatment presents the greatest relapse risk (Berends et al., 2016), though research has revealed that the window of risk could extend out as far as two years after treatment (Berends et al., 2018) and beyond (Strober et al., 1997).

These observed patterns suggest an illness episodic in nature, though the true trajectory of AN remains vague due to inconsistent definitions of "recovery" and "relapse." A clear understanding of the course of AN would be aided by the use of comprehensive and standardized definitions of these constructs. Further, accurate rates of relapse can only be extracted from a population deemed "recovered" using a multi-dimensional and empirically validated definition of recovery, as a discrete period of "illness" can only follow from a discrete period of "wellness." AN is a complex psychiatric disorder implicating pathology in cognitive, behavioral, and physical domains. Marked improvement or deterioration in only two of these domains—physical (i.e., weight status) and behavioral (i.e., calorically restricting, purging, etc.)—is often used to define recovery or relapse without consideration of the cognitive domain. Bardone-Cone and colleagues (2010) have made efforts to clarify the course of AN by proposing and implementing a comprehensive definition of recovery that relies on improvement across all three

symptom domains. In this paper we seek to further illuminate illness trajectories in AN by exploring and validating definitions of relapse in a population of individuals who have achieved recovery from AN per Bardone-Cone and colleagues' (2010) criteria.

Considering the chronicity of illness and high risk of death associated with AN, it is crucial to establish a clinically-meaningful definition of relapse for use as an indicator of intervention need in the highly sensitive period following treatment and recovery. Such a standardized definition would further our understanding of AN illness trajectories and allow for comparison of relapse rates between research studies. In addition to establishing a common set of criteria the field can use across clinical and research settings, clearly defining relapse in AN would allow us to demarcate a constellation of symptoms for which we can identify predictors; doing so will allow clinicians to distinguish those who might be at greater risk for a recurrence of AN symptoms after successful treatment and who might benefit from supplemental intervention or relapse prevention efforts. Further, exploring these definitions will elucidate *stages* of relapse to differentiate those who are "slipping" into illness in different symptom domains from those who are "fully" relapsed. Our team acknowledges that, in clinical practice, a standard definition of relapse should not be rigidly applied across patient cases. Skilled clinicians approach diagnosis and illness status with flexibility, using an individual's unique illness presentation, weight and symptom histories, and comorbid psychopathology to arrive at treatment decisions. However, validated definitions of recovery and relapse could help establish clinically-useful cutoff scores on widely used measures to indicate illness status, as exists for other psychiatric disorders such as depression (Zimmerman et al., 2014), and inform clinicians' recommendations.

In the past few years, research has confronted the issue of inconsistent definitions of relapse in AN. Khalsa and colleagues (2017) conducted a comprehensive literature review in an

effort to inventory existing definitions and rates of relapse in AN studies. Their work revealed that definitions of relapse varied considerably by weight loss metric (e.g., body mass index [BMI], ideal body weight [IBW], percentage of weight loss), by weight status cutoff (e.g., BMI below 16.5, BMI below 17.5, less than 85% IBW), by duration of returned symptoms (e.g., 6 to 8 weeks, 12 weeks, 3 months), and by the use of other indices of illness (e.g., Eating Disorder Examination [EDE] scores, "a need for psychiatric intervention"). These diverse definitions, unsurprisingly, yielded varied relapse rates ranging from 9% to 51% (Khalsa et al., 2017). Their review stressed the importance of establishing a clear definition of illness constructs in AN and proposed comprehensive definitions of recovery, remission, and relapse to use for future examination of AN symptom trajectories.

The present study seeks to extend Khalsa and colleagues' (2017) theoretical work by comparing definitions of relapse in a sample of individuals who have been deemed "recovered" from AN using a comprehensive definition of recovery (Bardone-Cone et al., 2010). To this end, we examined several unique conceptualizations of relapse and, within each relapse definition, examined clinically significant variables related to eating disorder symptomology and closely-related psychiatric illness and functioning constructs. We explored definitions of relapse that may be both more or less sensitive to the presence of eating disorder symptomatology—these definitions range from full-threshold AN according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), which includes reaching an underweight status in addition to cognitive and behavioral disturbances, to the presence of *only* disordered eating cognitions or *only* disordered eating behaviors in the absence of other eating disorder symptoms.

Relapse Constructs Overview

We examined eight distinct relapse constructs (Table 1). Three of these definitions utilize DSM-5 diagnostic criteria—namely, for 1) AN, 2) subthreshold or "atypical" AN (typically captured under other specified feeding and eating disorders [OSFED] but separated for the purposes of analyses), and 3) the criteria for all other eating disorders (including bulimia nervosa [BN], binge eating disorder [BED], and other OSFED diagnoses). These diagnostic definitions are typically used across clinical settings to identify both illness onset and relapse. We then explored Khalsa and colleagues' (2017) definitions of full and partial relapse. Finally, we examined two definitions, theoretically related to prior work from our research team (Bardone-Cone et al., 2010), which isolate cognitive symptoms and behavioral symptoms of eating disorders as novel relapse constructs.

DSM-5 Diagnostic Criteria Relapse Constructs

We first examined the utility of the full-threshold diagnosis of AN as a relapse construct (for a list of DSM-5 diagnostic criteria, see Appendix). The use of these diagnostic criteria as a relapse definition is valuable because they contain no symptom duration requirement; therefore, this definition captures individuals who do not meet the symptom duration criterion of other relapse definitions (e.g., one month). It is also likely widely used as a definition for illness in routine clinical practice, such as in primary care facilities. An AN diagnosis further captures individuals who are unambiguously in need of intervention, though is limited in its conservatism—it only captures the sickest individuals whose symptoms have progressed to full-syndrome illness and who would be more challenging to treat effectively compared to those with fewer or less severe symptoms. It is limited by a lack of operationalization of all diagnostic criteria (which include weight status, fear of weight gain, and body image disturbance), lack of

specificity of weight-loss behavior frequency (e.g., restriction, purging, exercise), and lack of reference scores on standardized measures (e.g., the Eating Disorder Examination Questionnaire [EDE-Q]).

In addition to full-threshold AN, we explored the use of subthreshold AN as a relapse construct. The DSM-5 provides diagnostic criteria for subthreshold (or "atypical") AN under the OSFED category (for a list of diagnostic criteria, see Appendix). A diagnosis of atypical AN requires that all of the criteria for full-threshold AN are met, except that the individual's weight is within or above the normal range despite significant weight loss (American Psychiatric Association, 2013). This diagnosis identifies those who engage in significant weight-loss behaviors and who may be on a dangerous weight loss trajectory but who have not yet reached an underweight status. However, the diagnostic criteria are vague, as there is no specified amount of weight one must lose to qualify for the diagnosis (Moskowitz & Weiselberg, 2017). As a relapse definition, the strength of this construct is that it captures individuals who may be in the early stages of relapse and who have not yet reached a clinically underweight status. As with DSM-5 full-threshold AN, this relapse definition is limited by its lack of operationalization of important constructs, lack of specificity of weight-loss behavior frequency, and lack of reference scores on standardized measures.

We also explored the strength and utility of the DSM-5 diagnostic criteria for other eating disorder diagnoses, including BN, BED, and OSFED (including subthreshold BN, subthreshold BED, and purging disorder, though excluding subthreshold AN as previously noted; for a list of diagnostic criteria, see Appendix) as relapse constructs. Research reveals a high likelihood of "diagnostic migration" to symptoms of BN and OSFED in individuals in recovery from AN (Eddy et al., 2008; Keel, et al., 2005; Schaumberg et al., 2019). For example, a Swedish register

study revealed that after a period of "recovery" (authors did not specify recovery criteria) from AN, 18% relapsed to eating disorders not otherwise specified (EDNOS, the DSM-IV equivalent of OSFED), 11% returned to AN, 2% relapsed to BN, <1% relapsed to BED, and 68% maintained a recovered status (Schaumberg et al., 2019). Therefore, by using these DSM-5 diagnostic categories as relapse constructs, we hope to capture those who might relapse into other eating symptom profiles after they achieved full recovery status from AN. One limitation of this category is the inconsistent definition of OSFED—there is not a clear consensus in the field on what meets criteria for this diagnosis and there is great diversity in OSFED diagnoses across research and clinical settings. A strength of these diagnostic categories as relapse constructs is that OSFED diagnoses identify individuals with subthreshold eating disorder presentations and those who might "migrate" to non-AN disordered eating symptom profiles after recovery, who might otherwise be missed if clinicians are only screening for underweight status, caloric restriction, overexercise, and weight loss.

Khalsa et al. (2017) Relapse Constructs

Khalsa and colleagues' (2017) proposed definitions of relapse utilize physical (e.g., weight status), cognitive, and behavioral dimensions. They distinguish between two relapse categories—full relapse and partial relapse—according to the duration of returned symptoms (three months of symptoms for full relapse and one month of symptoms for partial relapse) and the severity threshold of disordered eating behaviors (i.e., "significant" restricting, bingeing, or purging for full relapse and no such signifier in partial relapse). They suggest clear BMI thresholds and a means for operationalizing the cognitive component of relapse with an EDE score of $\geq$ 2 SDs of norms (Khalsa et al., 2017). A strength of these definitions is that a one-month "partial relapse" construct allows for a brief return of symptoms that may represent a

"slip" before individuals recover again, However, both the partial and full relapse definitions are limited in that they require underweight status at the level often used in full-threshold AN as well as a high EDE-Q score (i.e., $\geq$ 2 SDs approximates the 99th percentile of community norms; Mond et al., 2006; Luce et al., 2008; Welch et al., 2011; Aardoom et al., 2012), which will only capture those on the sicker end of the relapse spectrum.

Cognitive and Behavioral Relapse Constructs

Finally, we explored two novel definitions of relapse by separately examining the absence of cognitive recovery and behavioral recovery of AN, following a period of full recovery. While Khalsa and colleagues (2017) separated the relapse construct into "partial" and "full," depending on the duration and severity of a constellation of symptoms, we will separate relapse into symptom components. We will not examine the combination of cognitive and behavioral symptoms (in the absence of low weight status) as a relapse construct, because this symptom combination is captured within other diagnostic constructs (i.e., BN and OSFED DSM-5 relapse categories).

The cognitive relapse definition is founded on prior work on recovery by our research team (Bardone-Cone et al., 2010), which revealed similarities between individuals with full-threshold eating disorders (including AN) and a group of individuals who were otherwise considered recovered from an eating disorder (including cases of recovery from AN) along physical and behavioral dimensions but who still endorsed pathological attitudes towards body image, eating, and food. Bardone-Cone and colleagues (2010) referred to this group of individuals as being in "partial" recovery. This research suggested that the presence of eating disorder cognitions *alone* might be an indicator of clinically significant illness. This relapse construct is further inspired by evidence that lingering cognitive symptoms increase risk for

relapse in individuals with a history of eating disorders (Keel et al., 2005). Just as a core cognitive construct such as body image disturbance might be the "final hurdle" in recovery (Bardone-Cone, 2010), it might also be the "first to slip" in relapse—whether these symptoms suggest a step toward recovery or a step toward relapse depends on the discrete period of illness or wellness that came before. Including a broad definition of relapse to capture the re-emergence of these cognitive symptoms may therefore identify individuals early and direct them to treatment before their symptoms progress to full-threshold illness.

A strength of this cognitive relapse construct is that it captures transdiagnostic eating pathology that could progress to either an AN, BN, or BED symptom profile, as it only relies on distorted shape-, weight-, and eating-related cognitions and does not require weight loss or restriction, bingeing, or purging behaviors that might emerge later. It is also a prevention-oriented construct—because it is highly sensitive and requires only cognitive symptoms, rather than physical and behavioral symptoms, be present, it would identify individuals who might be "slipping" into worsening illness. One possible limitation of this definition, however, is that it might be overly inclusive and identify individuals who have normative body image concerns but who are otherwise able to maintain a true recovery status over time.

Finally, the behavioral relapse definition reflects the re-emergence of disordered eating behaviors in the absence of cognitive symptoms or low weight status in individuals who have previously reached recovery from an eating disorder. After the pathological eating behaviors have been eliminated, or "extinguished," in treatment, these behaviors could theoretically return via several classical and operant learning mechanisms (i.e., renewal, spontaneous recovery, resurgence, reinstatement, or rapid reacquisition; Bouton, 2011). For example, an individual who engaged in dieting, binge eating, and purging while away at college might receive treatment and

recover from an eating disorder while at home for the summer. When they return to college in the fall, they might experience a reflexive, automatic return to binge eating when exposed to the contextual environmental cues of the cafeteria dining hall even in the absence of disordered eating cognitions. Symptoms may also recur because they are precipitated by a worsening of the illness maintenance mechanisms that perpetuated these symptoms during the illness episode—according to Fairburn's widely-accepted cognitive behavioral formulation, these factors might include major life stressors, difficulty tolerating negative moods, clinical perfectionism, and low self-esteem (Fairburn et al., 2003).

If this relapse definition emerges as a robust indicator of illness, it might inform behavior-based intervention or relapse prevention strategies that target conditioning mechanisms. An advantage of this definition is that it is transdiagnostic and would capture individuals across AN, BN, BED, and OSFED symptom profiles. Another strength is that it is highly sensitive and inclusive of individuals who experience only minimal symptoms, thereby acting as an early identification tool (e.g., identifying a "lapse" before a "relapse"). Relatedly, however, a limitation is that, depending on the threshold of behavior frequency (e.g., once versus once a week for three months), it might be overly sensitive and identify individuals who had a one-time behavioral "slip" and are not truly in need of intervention. Moreover, eating disorder behaviors occur even in the general population (Hilbert et al., 2012), suggesting blurred boundaries between clinically significant and non-clinically significant behaviors.

Eating Disorder Constructs Potentially Related to Relapse

Disordered Eating Attitudes and Behaviors

We used several clinically relevant eating- and weight-related variables to validate relapse status, including measures of disordered eating thoughts and behaviors, exercise (i.e.,

motivation for and frequency of exercise), and measures of body-related cognitions (i.e., body shame, body surveillance, thinness expectancies). We expected scores on these measures to be higher in those who are considered to have relapsed across relapse definitions. Importantly, however, for the cognitive and behavioral relapse definitions, we expect that only scores on the cognitive measures (e.g., body shame and surveillance) and behavioral measures (e.g., compulsive exercise questions) will be elevated in the relapse groups, respectively, as the absence of symptoms in the other symptom domains are the defining features of these definitions.

Non-Eating Disorder Constructs Potentially Related to Relapse

Positive Eating- and Body-Related Attitudes

We also examined constructs representing positive eating- and body-related attitudes, such as body appreciation and intuitive eating. Body appreciation represents positive opinions about one's body regardless of physical appearance, acceptance of body despite weight, shape, and imperfections, respecting the body's needs, engaging in healthy behaviors, and protecting the body by rejecting unrealistic media-based expectations (Avalos et al., 2005). Body appreciation has been found to be significantly negative correlated with eating disorder symptomatology (Tylka & Kroon Van Diest, 2013). Intuitive eating refers to relying on physiological hunger and fullness cues to guide eating behaviors (Denny et al., 2013) and has also been found to be negatively correlated with eating disorder symptomatology (Bruce & Ricciardelli, 2016; Denny et al., 2013), positively correlated with psychological well-being (Tylka & Wilcox, 2006), and associated with positive treatment outcomes for AN (Richards et al., 2017). We expected scores on both the body appreciation and intuitive eating measures to be lower in those who are considered to have relapsed across all relapse constructs.

Non-Eating Disorder Psychopathology

In addition to these eating- and weight-related variables, we explored the relationship between relapse status and other correlates of general psychological illness. Individuals with AN have high lifetime rates of major depression and anxiety disorders (Halmi et al., 1990), and there is evidence that these comorbid disorders might wax and wane in severity depending on an individual's eating disorder recovery status. For example, there is evidence that depression symptoms improve with weight restoration in AN in a dose-response fashion (such that those at 80% IBW have lower depressive symptomatology than at intake, and again at 90% IBW have lower depressive symptomatology than at 80% IBW; Meehan et al., 2006). Similarly, using our group's definition of recovery from an eating disorder, Bardone-Cone (2010) found that a group of individuals fully-recovered from eating disorders had rates of current mood disorders comparable to individuals with no eating disorder history, suggesting that depressive symptoms might resolve among those who are fully recovered and no longer experience the cognitive symptoms of an eating disorder. Researchers have found, however, that symptoms of depression might reemerge in individuals with AN as weight is further restored or as they approach their ideal body weight (Mischoulon et al., 2011). A return of depressive symptoms in individuals whose depression previously remitted might therefore be an indicator of illness in individuals who no longer meet full DSM-5 diagnostic criteria for AN but who experience dissonance between their cognitions (e.g., desire for thinness) and behavior (e.g., eating appropriate calories). Accordingly, from a relapse perspective, the experience of depressive symptoms within a recovered sample might indicate a return to, or be a precursor of, eating pathology.

In contrast, while depressive symptoms might fluctuate alongside AN symptomatology, evidence suggests that symptoms of anxiety disorders may persist after recovery from an eating

disorder. Bardone-Cone and colleagues (2010) found that groups of fully and partially recovered individuals still had higher rates of anxiety disorders compared to controls (though lower rates compared to those with an eating disorder). Obsessionality, or the preoccupation with specific thoughts or acts (Denys, 2011), may persist through AN weight restoration and recovery similar to the pattern observed with anxiety (Holtkamp et al., 2005). There is evidence that obsessionality may decrease as weight improves (Pollice et al., 1997). In the present study, therefore, we anticipated that depression, anxiety, and obsessionality will be higher in those who are considered to have relapsed across all relapse constructs.

Personality Traits

We also examined the relationship of personality characteristics, including negative urgency and low distress tolerance, to eating disorder relapse status. Broadly, impulsivity has been associated with poorer outcomes in AN (Fichter et al., 2006). One facet of impulsivity, negative urgency, or the tendency to act impulsively in response to negative affect (Riley et al., 2015), may be a particularly salient impulsivity dimension for eating disorder recovery, especially for those with a history of binge eating and/or purging (i.e., those with the binge-purge subtype of AN; Bardone-Cone et al., 2016). Prior work from our team revealed that individuals fully recovered from an eating disorder experienced significantly less negative urgency than those with a current eating disorder and were comparable to controls (Bardone-Cone et al., 2016). Additionally, distress tolerance, or the ability to withstand and accept the experience of negative affect, is impaired among individuals with AN; Hambrook and colleagues (2011) found that AN participants scored higher than controls on a measure of difficulties with distress tolerance, and greater eating pathology in the AN group was associated with greater maladaptive beliefs about both expressing and experiencing emotions. Consistent with these findings, we

anticipated greater negative urgency and lower distress tolerance in individuals who meet criteria within the relapse definitions.

Psychosocial Functioning and Quality of Life

Finally, we examined markers of psychosocial functioning and quality of life, including social support, loneliness, and perceived stress, as indicators of relapse. Female patients with AN may be less inclined to seek social support compared to healthy individuals (Villa et al., 2009); one study found that individuals with a history of an eating disorder described feeling disconnected from social networks because of shame related to their eating and weight loss habits and fear that others would not understand or support them (Linville et al., 2012). Relatedly, individuals may report feelings of loneliness from perceived social isolation, even if they are not physically isolated from others, because of eating-related secrecy (Pritchard & Yalch, 2009). However, there is evidence that interpersonal relationships and feelings of loneliness may improve with recovery (Harney et al., 2014). Low social support and self-reported loneliness in a sample of recovered individuals, therefore, might indicate social isolation and feelings of disconnection from friends and family because of shame surrounding the return of disordered eating symptoms. Perceived stress, which reflects one's subjective appraisal of experiences and beliefs about one's ability to cope with them, has been associated with recovery from an eating disorder—Bardone-Cone and colleagues (2010), using our team's aforementioned recovery criteria, found that individuals in recovery scored similarly to individuals with no eating disorder history on levels of perceived stress and had lower levels than those in partial recovery or active illness. Taken together, we expected that social support will be lower and feelings of loneliness and perceived stress will be greater in individuals who are considered to have relapsed across all relapse definitions.

Study Aims and Hypotheses

Our study aims and hypothesis are theoretically influenced by those of Ackard and colleagues (2014), who examined a spectrum of constructs of eating disorder remission by comparing rates of remission by definition, levels of agreement between definitions, and the distinctions between definitions on quality of life variables. The authors did not offer hypotheses about which specific remission definition would emerge as the best indicator of wellness. Rather, they stated that the most robust remission definition would be in moderate to high agreement with other definitions and would reveal higher disease-specific and generic quality of life ratings in those identified as remitted (Ackard et al., 2014). Accordingly, in the present study we offer three study aims as well as exploratory hypotheses concerning the most robust indicator of relapse in individuals recovered from AN.

Aim 1: We compared rates of relapse across relapse conditions.

Hypothesis 1: We hypothesized that the highest rates of relapse would be observed using the Behavioral Relapse (Any Frequency) and Cognitive Relapse definitions and the lowest rates of relapse would be observed using the DSM-5 diagnostic criteria for AN.

Aim 2: We explored rates of agreement between relapse conditions.

Hypothesis 2: We hypothesized that a robust definition of relapse would demonstrate high agreement (Cohen's Kappa $\geq$ 0.6; Viera & Garrett, 2005) with other theoretically related relapse constructs. Importantly, because some of the relapse constructs are, by definition, mutually exclusive based on weight status (e.g., Cognitive and Behavioral Relapse constructs require BMI $\geq$ 18.5 while Khalsa et al. [2017] definitions require BMI $\leq$ 18.5), we expected high agreement between the relapse constructs specifying underweight status (i.e., DSM-5 AN, Khalsa full relapse, and Khalsa partial relapse definitions) and high agreement between the relapse

constructs specifying not-underweight status (i.e., DSM-5 subthreshold AN, non-AN eating disorders, Cognitive Relapse, and Behavioral Relapse definitions). We did not expect high agreement between the Cognitive and Behavioral Relapse definitions, as these are each defined by the absence of symptoms in the other definition.

Aim 3: We determined which definitions are sensitive enough to differentiate relapsed individuals from non-relapsed individuals on relevant variables.

Hypothesis 3: We hypothesized that a robust definition of relapse would yield statistically and clinically significant differences between relapsed and non-relapsed individuals on variables related to disordered eating attitudes and behaviors, positive eating- and body-related attitudes, non-eating disorder psychopathology, personality traits, and psychosocial functioning and quality of life.

Methods

Participants and Recruitment

Participants included 223 women with a history of an eating disorder recruited from former patients (18 or older) seen at eating disorder treatment centers (University of North Carolina at Chapel Hill [UNC] Center of Excellence for Eating Disorders, Duke Center for Eating Disorders). Individuals recruited through eating disorder centers had been treated for an eating disorder and received letters of invitation and follow-up phone calls. To be eligible for the study, all participants must have been out of intensive treatment for at least one year. All interested individuals were screened via phone for lifetime eating disorder history using the eating disorder module of the Structured Clinical Interview for DSM-IV (SCID; First et al., 2002) with DSM-5 criteria applied to ensure that they met diagnostic criteria for a lifetime history of an eating disorder (AN, BN, BED, or OSFED).

This was a longitudinal study involving data collection at three time points, each about six months apart (see Figure 1). Of the total 223 individuals who participated at baseline/Time 1, 210 (94%) also participated in the data collection about six months later (Time 2), and 187 (84% of T1 group) participated about 12 months after baseline (Time 3).

Study Procedures

At baseline/Time 1 (T1), participants completed an online survey remotely (e.g., at home). This survey included an array of questionnaires covering disordered eating and body image among others constructs. Within a month of having completed the survey, participants came to the UNC campus for a 5-6 hour visit, which included a neurocognitive battery, a set of diagnostic interviews among other interviews, the measurement of height, weight, and body composition, and a blood draw for examining biomarkers. At Time 2 (T2) the same T1 online survey was completed remotely, followed up by a phone interview including diagnostic interviews and other interviews. Time 3 (T3) included the same study components as in T1; the visit at T3 took about 3-4 hours since the focus diagnostically and for other interviews was the past six months instead of lifetime. A small number of participants were unavailable for an in-person visit at T3 and completed the assessments that could be done remotely (e.g., interviews) over the phone (n=8). Participants were compensated financially for participating in each of the time points. The Institutional Review Boards of UNC and Duke University approved this study.

Eating Disorder Recovery Status Definition and Assessment Tools

Analyses for the current study were conducted on those with a lifetime history of AN who completed all three study timepoints and who, at T1, were identified as having reached full recovery from an eating disorder per Bardone-Cone and colleagues' (2010) definition: absence of an eating disorder diagnosis; physical recovery, operationalized as a BMI $\geq$ 18.5 kg/m^2;

behavioral recovery, operationalized as no binge eating, vomiting, laxatives, or fasting in the past three months; and cognitive recovery, operationalized as all four EDE-Q subscales within 1 SD of age-matched community norms (Mond et al., 2006). See Table 3.

The SCID (First et al., 1995) was used to assess lifetime and current eating disorder diagnoses at T1. The interview was modified slightly to make possible DSM-5 diagnoses for AN, BN, BED, and OSFED. Study staff administering SCIDs consisted of post-baccalaureate research assistants and doctoral clinical psychology students who were trained on DSM eating disorders diagnostic criteria, observed the primary investigator (A. B.-C.) administering several SCID assessments, and conducted role plays. All staff members were approved to conduct independent assessments by the primary investigator. For the physical recovery domain, BMI was calculated from measured weight and height at T1 using a digital scale and stadiometer. A BMI of at least 18.5 kg/m^2 was required, per the World Health Organization's recommendation of 18.5 as the low end of "normal" weight (Björntorp, 2002). For the behavioral recovery domain, the presence of any objective binge eating, vomiting, laxative use, or fasting over the past three months was assessed at the T1 visit using annotated calendars. The absence of all four eating disorder behaviors across the past three months was required in order to meet criteria for behavioral recovery. Finally, the cognitive recovery domain was assessed with the Eating Disorder Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994), which was administered at T1 in the online survey and contains four subscales broadly covering eating disorder cognitions over the past 28 days: Restraint, Eating Concern, Weight Concern, and Shape Concern. Bardone-Cone et al. (2010) proposed using scoring within 1 SD of age-matched community norms for each of the EDE-Q subscales as an indicator of cognitive recovery (Mond et al., 2006).

Operationalization of Relapse Definition Components

The DSM-5 diagnostic criteria relapse constructs were operationalized using the DSM-5 criteria for each disorder (see Appendix).

As previously noted (see Table 1), Khalsa and colleagues' (2017) definitions utilize physical (i.e., BMI), behavioral, and cognitive symptoms, with duration of returned symptoms and frequency of disordered eating behavior distinguishing between partial relapse (1 month's duration, any presence of disordered eating behavior) and full relapse (3 months' duration, significant disordered eating behavior present). In the case of an individual's symptoms returning for only part of one month, the authors note that, from their clinical perspective, the presence of symptoms for the majority of the one-month assessment period still qualifies as partial relapse, even if these symptoms are not sustained across all domains for the full month (S. Khalsa, personal communication, April 1, 2019).

To be considered in either full *or* partial relapse, Khalsa et al. (2017) specify that individuals should be underweight (defined by Khalsa and colleagues as BMI less than or equal to 18.5 or less than or equal to 85% IBW). For full relapse, they specify that individuals are engaging in "significant" restricting, bingeing, or purging, though for partial relapse they do not include the "significant" specifier. Upon further inquiry, Khalsa and colleagues noted that "significant" disordered eating behaviors would include a regular behavioral pattern, a return to prior patterns of disordered eating, or would be a new manifestation (i.e., an "evolution or transformation") of the eating disorder (S. Khalsa, personal communication, April 1, 2019). For the purposes of the current study and in the absence of any established or meaningful behavioral frequency threshold, we defined "significant" using the frequency of at least once weekly for the past three months (a frequency criterion that is used for the DSM-5 BN diagnosis); thus, meeting

this frequency threshold for behavior would warrant consideration for full relapse. We considered individuals to have met the behavioral criterion for partial relapse if they engaged in any disordered eating behaviors, even if only once, over the prior month. Khalsa and colleagues (2017) further suggest that individuals in either full or partial relapse endorse clinically significant eating disorder symptomatology as evaluated by an EDE global score exceeding 2 SDs of community norms. They suggest using the global score over the subscale scores because the global score gives a more general impression of an individual's disordered eating thoughts (S. Khalsa, personal communication, April 1, 2019). In the current dataset, we are able to examine the frequency of behaviors over the past month and past three months (using the aforementioned annotated calendars), though we are limited by the timeframe assessed by the EDE-Q (i.e., past month) and by BMI measurement occurring only at each study time point and not tracked across the 1-month or 3- month time periods proposed by Khalsa and colleagues (2017).

The cognitive relapse construct defines relapse as endorsing clinically significant disturbance in body image or pathological concerns about eating (as evaluated by EDE-Q global scores greater than 1 SD of community norms; Mond et al., 2006) among individuals who have not reached an underweight status (BMI above or equal to 18.5) and who have not been engaging in pathological eating behaviors (i.e., fasting, bingeing, or purging) in the past three months. Our use of the BMI cutoff in lieu of the DSM-suggested 85% IBW cutoff is based on research from Thomas and colleagues (2009), whose work revealed significant discrepancies in the methods used to calculate ideal and expected body weight as well as the associated rates of diagnosis. These authors and others (Brown et al., 2014) recommend the use of an 18.5 BMI cutoff, as the BMI algorithm can be applied to any height, has a standard formula, and is not vulnerable to the

changes in population weight trends over time like the Metropolitan Life tables used for IBW (Thomas et al., 2009).

The behavioral relapse construct defines relapse as engaging in pathological eating behaviors (i.e., fasting, bingeing, or purging) over the past three months among individuals who have not reached an underweight status (BMI above or equal to 18.5) and who do not endorse clinically significant disturbance in body image or pathological concerns about eating (as evaluated by EDE-Q subscale scores within one SD of community norms). For the behavioral relapse definition we will examine two different behavior frequencies: 1) any presence of pathological behaviors in the past three months, even if it occurred only once, and 2) any pathological behaviors that occurred, on average, at least once weekly for the past three months, a frequency criterion that is used for the DSM-5 BN diagnosis.

Assessment Tools Used to Measure Relapse Definition Components

A summary of the eight relapse definitions under consideration is presented in Table 1. For the relapse definitions derived from diagnostic criteria, we used the SCID to determine presence of the following DSM-5 eating disorder diagnoses at T2 and T3: AN, Subthreshold AN, and the non-AN-related eating disorder diagnoses of BN, BED, and OSFED (which includes OSFED-subthreshold BN, OSFED-subthreshold BED, and OSFED-purging disorder).

For the physical criteria of Khalsa and colleagues' (2017) full and partial relapse definitions and the cognitive relapse and behavioral relapse definitions, we used measured weight and height (digital scale and stadiometer) to compute BMI at T3 and self-reported weight and height for BMI at T2.

For the behavioral criteria of Khalsa and colleagues' (2017) full and partial relapse definitions and the cognitive relapse and behavioral relapse definitions, study staff interviewed

participants using annotated calendars assessing the frequency of binge eating, purging (e.g., via vomiting, laxative, or diuretic use), or fasting (defined as "intentionally [and not for religious reasons] going without eating for a 24-hour period") over the past three months, with only the past 1-month period used for Khalsa and colleagues' partial relapse definition.

For the cognitive criteria of Khalsa and colleagues' (2017) full and partial relapse definitions and the cognitive relapse and behavioral relapse definitions we used participants' global scores on the EDE-Q. Although Khalsa et al. (2017) suggest using the EDE, they acknowledge that the EDE-Q may be easier and less time intensive to administer (S. Khalsa, personal communication, April 1, 2019). In the current study we used the global score on the EDE-Q, the self-report version of the EDE, as a proxy for the EDE. The self-report EDE-Q has generally demonstrated high correlation with the clinician-administered EDE, though EDE-Q subscale scores tend to be higher (Mond et al., 2004). As previously noted, the EDE-Q only assesses symptoms over the past month.

Measures for Assessing Validity of Relapse Definitions

See Table 2 for a list of the measures used for assessing validity of the definitions by comparing relapsed and non-relapsed individuals on variables related to disordered eating attitudes and behaviors, positive eating- and body-related attitudes, non-eating disorder psychopathology, personality traits, and psychosocial functioning and quality of life. Information about coefficient alphas for these measures at T2 and T3 are included in this table.

Disordered Eating Attitudes and Behaviors

Broad disordered eating behaviors and attitudes were assessed using the Eating Attitudes Test-26 (EAT-26; Garner et al., 1982), a 26-item self-report measure consisting of questions about eating behaviors ("I avoid eating when I am hungry") and attitudes ("I feel that food

controls my life") that participants rate on a scale from 1 (*never*) to 6 (*always*). Items rated as 1, 2, or 3 are given a score of "0," while items rated 4, 5, or 6 are scored as "1," "2," or "3," respectively. Higher scores on the EAT-26 are related to greater levels of disordered eating, and a cutoff score of 20 or above signifies a probable eating disorder (King, 1989, 1991). The measure has demonstrated good internal consistency and test-retest reliability in samples of young women (Carter & Moss, 1984; Garner et al., 1982).

Obsessional eating disorders thoughts were measured using a set of eight questions designed by our research team, most of which were reported in Bardone-Cone et al. (2010). The questions assessed the amount of waking life devoted to thinking about weight, shape, exercise, and eating, rated on a scale from 1 (*no time or almost no time*) to 5 (*almost all the time or all the time*), as well as the ease with which these thoughts can be stopped, rated on a scale from 1 (*extremely easy*) to 5 (*extremely difficult*). Higher scores on this measure indicate greater levels of obsessional thinking.

Motivation for exercise was assessed using the Reasons for Exercise Inventory (REI; Silberstein et al., 1988). Respondents rated the personal importance of 24 possible exercise motivators (e.g., "To be slim," "To meet new people," "To cope with stress, anxiety") on a scale from 1 (*not at all important*) to 7 (*extremely important*). Prior factor analytic work with the REI (Cash et al., 1994) has revealed four meaningful factors of exercise motivation: for appearance/weight management (e.g., exercising to improve physical appearance), for fitness/health management (e.g., exercising to improve cardiovascular health), for stress/mood management (e.g., exercising to cope with stress), and for socializing (e.g., exercising to spend time with friends). In their original analysis using a non-clinical sample of 100 healthy exercising women, the authors found that the items "To improve my muscle tone" and "To

maintain my current weight" did not load onto any of the four factors. However, we retained these items for the present study, as we theorized that they would be related to the existing four factors in a sample of individuals in recovery from an eating disorder. For our analysis we included "To improve my muscle tone" in the Fitness/Health Management subscale and "To maintain my current weight" in the Appearance/Weight Management subscale. Individuals were instructed to skip the section if they do not engage in exercise—correspondingly, two individuals skipped the measure at T2 and two individuals skipped the measure at T3. The individuals who were missing items for this measure did not meet criteria for any relapse definition at either timepoint and were therefore categorized as "Not Relapsed" for analyses. In a sample of undergraduate women, coefficient alphas for these factors ranged from .73 to .91 (Cash et al., 1994).

We assessed frequency of and attitudes towards exercise in several ways. First, we asked how many hours of cardiovascular exercise (e.g., running, cycling) and how many hours of strength/resistance exercise (e.g., yoga, weight-lifting) participants did in a typical week. Next, participants rated eight questions on a sliding visual analog scale about their exercise-related cognitions (e.g., "Do you feel guilty that you have somehow 'let yourself down' when you miss your exercise session?") and habits (e.g., "Do you continue to exercise even when you have sustained an exercise-related injury?"). These items capture the concept of compulsive exercise, with many of these concepts represented in the expert clinical conceptualization of unhealthy exercise derived from Delphi methodology (Noetel et al., 2017). For these questions, higher numbers indicate a greater presence of pathological exercise cognitions and behaviors.

Body shame was measured by the 8-item Body Shame subscale of the Objectified Body Consciousness Scale (OBCS; McKinley & Hyde, 1996). The Body Shame subscale assesses

feelings of shame when a respondent feels their body does not confirm to cultural standards of beauty ("I feel like I must be a bad person when I don't look as good as I could"). Items on the measure are rated on a 7-point scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). Higher scores on the Body Shame subscale indicate greater feelings of body shame. The authors of the scale reported an internal consistency alpha coefficient of .75 and report that the OBCS overall demonstrates validity in young and middle-aged women (McKinley & Hyde, 1996).

Body surveillance was measured by the 8-tem Body Surveillance subscale of the OBCS (McKinley & Hyde, 1996). This subscale assesses the extent to which respondents evaluate their own body as if they were an outside observer ("During the day, I think about how I look many times"). As with the Body Shame subscale, items are rated on a 7-point scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*), with higher scores indicating greater levels of body surveillance. The authors of the scale reported an internal consistency alpha coefficient of .89 (McKinley & Hyde, 1996).

Thinness and restricting expectancies were assessed using a brief version of the Thinness and Restricting Expectancy Inventory (TREI; Hohlstein et al., 1998). The TREI consists of eight self-report items about thinness- and caloric restriction-related cognitions ("I would be more good looking if I were thin"), which respondents rate on a scale from 1 (*Completely Disagree*) to 7 (*Completely Agree*). The brief version of this measure has demonstrated validity and reliability in female samples (Bardone-Cone et al., 2010).

Positive Eating- and Body-Related Attitudes

Body appreciation was measured using the Body Appreciation Scale (BAS; Avalos et al., 2005). The BAS consists of 13 self-report items regarding body appreciation ("I respect my body," "Despite its flaws, I accept my body for what it is") that respondents rate on a scale from

1 (*Never*) to 5 (*Always*). The BAS has demonstrated internal consistency and stability of scores as well as construct and incremental validity in a sample of women (Avalos et al., 2005).

Intuitive eating was measured using the Intuitive Eating Scale (IES; Tylka, 2006). The IES is a self-report measure consisting of 21 items assessing intuitive eating behaviors ("When I'm eating, I can tell when I'm getting full," "I trust my body to tell me when to eat") that respondents rate on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). The IES has demonstrated validity, internal consistency, and test-retest reliability in a sample of college women (Tylka, 2006).

Non-Eating Disorder Psychopathology (Depression and Anxiety)

The Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) was used to assess depressive symptomology. The CES-D consists of 20 self-report items assessing depressive symptoms over the prior week ("I felt that I could not shake off the blues even with help from my family or friends"), which respondents rate on a frequency scale from 1 (*rarely or none of the time, less than 1 day*) to 4 (*most or all of the time, 5 to 7 days*). Higher scores on the CES-D indicate greater depressive symptomology. This measure has been found to have high internal consistency, adequate test-retest reliability, and construct validity within a general population (Radloff, 1977).

Trait anxiety was measured using the trait anxiety scale of the Spielberger State-Trait Anxiety Inventory (STAI; Spielberger et al., 1970), a 20-item self-report measure assessing an individual's general tendency to experience anxiety. Respondents rate statements about anxiety ("I am calm, cool, and collected," "I worry too much over something that really doesn't matter") on a scale from 1 (*Almost Never*) to 4 (*Almost Always*). The STAI has demonstrated construct

validity (Hedberg, 1972), convergent and divergent validity (Watson & Clark, 1984; Taylor et al., 1992), and test-retest reliability (Metzger, 1976; Rule & Traver, 1983).

Anxiety symptoms were further assessed using the Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010), a 20-item self-report questionnaire assessing four symptom dimensions of obsessive-compulsive disorder: fear of germs/contamination; fear of being responsible for harm, injury, or mistakes; intrusive and distressing unwanted thoughts; and a need for symmetry, completeness, or for things to be "just right." For each 5-item subscale, items are rated on a Likert scale from 0 to 5 to indicate the time spent on obsessions and rituals, avoidance behavior, associated distress, functional impairment, and difficulty disengaging from the obsessions, with higher subscale and total scores indicating greater obsession- and compulsion-related pathology. The measure also generates a total score. The DOCS has demonstrated excellent internal consistency across clinical and non-clinical samples (Abramowitz et al., 2010).

Personality Traits

Distress tolerance was measured using the Distress Tolerance Scale (DTS; Simons & Gaher, 2005), a 15-item self-report measure consisting of a series of statements related to one's ability to tolerate emotional distress ("I can't handle feeling distressed or upset"), appraise distress ("My feelings of distress or being upset are not acceptable"), attend to (or avoid) negative emotions ("When I feel distressed or upset, all I can think about is how bad I feel"), and regulate negative affect ("When I feel distressed or upset, I must do something about it immediately"). Items are rated on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) with higher scores representing higher distress tolerance. The DTS has demonstrated convergent, discriminant, and criterion validity as well as test-retest reliability (Simons & Gaher, 2005).

The negative urgency subscale of the UPPS Impulsive Behavior Scale (UPPS), a 45-item self-report measure, was used to assess negative urgency, or the tendency to respond to negative affect with impulsive behavior. All items are scored on a scale from 1 (*Agree Strongly*) to 4 (*Disagree Strongly*). The UPPS has demonstrated reliability and validity (Anestis et al., 2007; Whiteside & Lynam, 2001; Whiteside et al., 2005).

Psychosocial Functioning and Quality of Life

Social support and loneliness were assessed using the emotional support items on the Social Health subscale of the Patient-Report Outcomes Measurement Information System (PROMIS; Castel et al., 2008). These items were designed to assess perceived feelings of being cared for and valued as a person and an individual's confidant relationships (HealthMeasures, 2018). The PROMIS emotional support items are 16 self-report questions evaluating the respondents' relationships ("I have someone who makes me feel appreciated," "I have someone I trust to talk with about my problems") rated on a scale from 1 (*Never*) to 5 (*Always*). Higher scores on this measure reflect a greater sense of social support. The PROMIS items have demonstrated reliability and construct validity (Cella et al., 2010).

Perceived stress was assessed using the Perceived Stress Scale (PSS; Cohen et al., 1983), a 14-item self-report questionnaire assessing the degree to which respondents appraise situations in their lives as stressful (e.g., "In the last month, how often have you felt confident about your ability to handle your personal problems?"). Items are rated on a Likert scale of 0 (*Never*) to 4 (*Often*) and higher scores indicate greater perceived life stress. The PSS has demonstrated good reliability as well as adequate concurrent and predictive validity (Cohen et al., 1983).

Data Analytic Plan

We examined relapse at the T2 and T3 assessment time points (refer to Figure 1 for procedure timeline). For each participant, we separately determined which, if any, relapse definition they met at T2 or T3. To examine differences in self-report measure scores between relapsed and non-relapsed individuals for Aim 3, we used the following procedure: 1) if an individual only met relapse criteria at T2, we examined T2 constructs as part of validating relapse status at T2; 2) if an individual only met relapse criteria at T3, we examined T3 constructs as part of validating relapse status at T3; 3) if an individual met relapse criteria at both T2 and T3, we focused on T2 constructs in validating relapse status at T2. Thus, for Aim 3, data from relapsed individuals came from the time point they *first* relapsed. The non-relapsed individuals in Aim 3 were those who did not relapse at all over the course of the study period; we used the mean value of their T2 and T3 scores on each measure for inclusion in Aim 3 analyses.

To examine Aim 1 (comparison of rates of relapse across relapse conditions), we calculated the percentage of the sample considered to have relapsed per definition at each time point.

To examine Aim 2 (rate of agreement between relapse conditions), we analyzed the rates of agreement between relapse definitions using Cohen's Kappa (Viera & Garrett, 2005) separately at T2 and T3.

To examine Aim 3 (differentiation of relapsed individuals from non-relapsed individuals on relevant variables as a way of evaluating validity of relapse definitions), we compared the mean scores on each measure (listed in Table 2) across relapse definitions between those who have relapsed and those who have not relapsed, according to each construct. As noted above, we used participants' self-report measures from the time point at which they *first* met relapse criteria

(i.e., at T2 or T3). Thus, if they met relapse criteria at both time points (T2 and T3), we used self-report measures from the time point at which they *first* met relapse criteria (i.e., T2). Also as previously noted, the comparison group for Aim 3 was those individuals who never relapsed across the study period. We used *t*-tests to compare means between groups for all continuous variables. On measures for which there are established, validated clinical cutoff scores (e.g., EAT-26, CES-D, DOCS), we additionally analyzed contingency tables with Fisher's exact test (due to small cell sizes) to determine whether individuals who were categorized as relapsed are overrepresented among those scoring within a clinical range. For the EAT-26, we used a cutoff score of 20 or higher (based on recommendations by authors Garner et al., 1982). For the CES-D, we used a cutoff score of 16 or higher (based on recommendations by Radloff & Locke, 1986; Radloff & Teri, 1986; Radloff, 1991). For the DOCS, we used a cutoff score of 18 (based on recommendations by Abramowitz et al., 2010). Because of the small sample size, we did not make corrections for multiple testing.

Results

Description of Sample

Our final sample for analyses was 26 individuals who met full criteria for recovery from an eating disorder at T1 (see Table 3 for recovery criteria), who had a lifetime history of AN, and who had completed all three timepoints. The sample was 100% female and 100% white-identified, with the vast majority of individuals identifying as non-Hispanic (96.2%, *n*=25). One participant identified as Hispanic (3.8%). Most of the participants were single (73.1%, *n*=19) or living with a partner (15.4%, *n*=4) while 7.7% were married (*n*=2) and 3.8% were divorced (*n*=1). Most of the sample identified as heterosexual (96.2%, *n*=25) with one participant identifying as bisexual (3.8%, *n*=1). Most of the sample were current full- or part-time students

(57.7%, *n*=15) and 42.3% were not a current student, but instead were a full- (26.9%, *n*=7) or part-time (7.7%, *n*=2) wage earner. One individual was unemployed (3.8%).

At T1, participants ranged in age from 19 to 45 years old (*M*=25.23, *SD*=7.41) and their BMIs, using measured height and weight, ranged from 18.73-29.09 (*M*=21.73, *SD*=2.41). At T2 participant BMIs ranged from 19.05 to 28.30 (*M*=21.80, *SD*=2.17) and at T3 BMIs ranged from 17.52 to 32.02 (*M*=22.06, *SD*=2.79). Only one participant had a BMI below 18.5 at T3.

Aim 1: Comparing Rates of Relapse Across Relapse Definitions

Out of 26 total participants, 21 individuals (80.8%) did not relapse at all over the course of the study and five individuals (19.2%) relapsed at either T2 or T3 (see Table 4). Relapse rates were examined separately at both timepoints (see Table 5). At T2, four total participants relapsed; three into the Behavioral Relapse (Any Frequency) definition (11.5% of total sample) and one into the Cognitive Relapse definition (3.8% of total sample). At T3, three total participants relapsed; one into the DSM-5 SubAN definition (3.8% of total sample), one into the DSM-5 Other Eating Disorders definition, namely OSFED-purging disorder (3.8% of total sample), and one into the Behavioral Relapse (Any Frequency) definition (3.8% of total sample). No participants met relapse criteria for DSM-5 AN, Khalsa Full Relapse, Khalsa Partial Relapse, or Behavioral Relapse (1x/Week Frequency) at any time during the study period.

One participant met relapse criteria for Behavioral Relapse (Any Frequency) at T2 and relapse criteria for DSM-5 SubAN at T3 (participant 144; see Table 4). One participant (participant 450) relapsed to the Behavioral Relapse (Any Frequency) category at T2 and then migrated to DSM-5 Other Eating Disorders (OSFED-Purging Disorder) at T3. Thus, of the two participants who met criteria for a DSM-5 eating disorder at T3, both exhibited behavioral prodromes of relapse at T2 (engagement in any bingeing, purging, or fasting over the past three

months), suggesting early indicators of relapse to meeting eating disorder diagnostic criteria in the future. The remaining three participants who relapsed during the study only met relapse criteria at *one* timepoint—participants 167 and 375 met criteria for Behavioral Relapse (Any Frequency) and Cognitive Relapse, respectively, at T2, and did not meet criteria for any relapse definition at T3. These two cases may best capture the idea of *lapses*—that is, temporary slips that may not develop into more clinically severe relapse conditions. Participant 132 did not meet criteria for any relapse definition at T2 but met for Behavioral Relapse (Any Frequency) at T3.

Aim 2: Exploring Rates of Agreement Between Relapse Conditions

We next examined rates of agreement between relapse conditions separately at T2 and T3 using Cohen's Kappa (Viera & Garrett, 2005). Due to the few relapse definitions met, we were only able to examine agreement between two relapse definitions at Time 2 (Behavioral Relapse-Any Frequency and Cognitive Relapse) and between three relapse definitions at T3 (DSM-5 SubAN, DSM-5 Other Eating Disorders, and Behavioral Relapse- Any Frequency). Furthermore, researchers (Bujang & Buharam, 2017) recommend a minimum sample size of two, and ideally a minimum sample size of 11, to examine Cohen's Kappa, though our sample size for agreement analyses was four at T2 and only two at T3. Therefore, it is unclear to us whether we can accurately interpret the Kappa coefficients. However, qualitatively, none of the relapse definitions had any individual overlap (i.e., in no cases did the same participant meet criteria for more than one definition at a time) (see Tables 6 & 7).

Aim 3: Comparing Relapsed vs. Not Relapsed Individuals' Scores on Clinically Relevant Measures

Finally, to address our third aim regarding comparisons on clinically-relevant measures, we conducted independent samples t-tests comparing individuals who never relapsed over the

study period (n=21) to individuals who met relapse definition criteria. Due to our general low sample size and low group sizes across relapse definitions, we combined individuals with SubAN and Other EDs (Purging Disorder) into one "DSM-5 ED" category at T3. Because so few participants met criteria for relapse definitions across the study period, we were only able to conduct pairwise mean comparisons for the Behavioral Relapse (Any Frequency) definition (n=4) and for the DSM-5 Eating Disorders definition (n=2).

Overall, only six out of 42 pairwise comparisons were statistically significant at $p < .05$ (see Table 8). As previously noted, we did not correct for multiple comparisons given the low sample size and, thus, low power. For the Behavioral Relapse (Any Frequency) definition, individuals who relapsed reported engaging in twice as many hours of cardio and strength training exercise (M=10.00, SD=3.74) relative to individuals who did not relapse (M=5.01, SD=3.17), $t(23)=2.81$, p=.010. They also had statistically significantly higher scores on the Delphi panel questions assessing the degree to which their exercise is compulsive (M=73.59, SD=9.38) relative to individuals who did not relapse (M=38.39, SD=23.48), $t(11.89)=5.07$, p<.001. Also related to exercise, individuals who relapsed according to the Behavioral Relapse (Any Frequency) definition reported a greater likelihood of engaging in exercise for stress management purposes (M=24.25, SD=0.96) relative to individuals who did not relapse (M=19.80, SD=5.47), $t(22.00)=3.38$, p=.003, and reported a greater likelihood of engaging in exercise for socializing purposes (M=9.75, SD=1.26) relative to individuals who never relapsed over the study period (M=6.08, SD=2.96), $t(21)=2.40$, p=.026.

For the DSM-5 Eating Disorders (i.e., SubAN and OSFED- Purging Disorder) category, those who met relapse definition criteria reported higher levels of depression on the CES-D (M=19.00, SD=5.66) than their non-relapsed counterparts (M=10.07, SD=5.75), $t(21)=2.10$,

p=.048. Similar to the pattern observed for the Behavioral Relapse (Any Frequency) group, individuals in the DSM-5 Eating Disorders relapse group endorsed engaging in a statistically significantly higher number of hours of exercise (*M*=12.00, *SD*=2.83) compared to those who did not relapse (*M*=5.01, *SD*=3.17), *t*(21)=2.99, *p*=.007.

Two pairwise comparisons approached significance (see Table 8). Individuals who relapsed into the DSM-5 Eating Disorders relapse group also reported higher rates of compelled or compulsive exercise (*M*=73.25, *SD*=8.13) relative to those who did not relapse (*M*=38.39, *SD*=23.48), *t*(21)=2.049, *p*=.053, as well as lower overall social support (*M*=49.00, *SD*=0.00) relative to those who did not relapse (*M*=64.86, *SD*=10.87), *t*(21)=-2.02, *p*=0.56. Other patterns emerged from t-tests at a trend level (i.e., $p > .05$ but $< .10$), including higher EAT-26 scores in both relapse Behavioral (Any Frequency) and DSM-5 Eating Disorders conditions relative to non-relapsed individuals (*p*=.086 and *p*=.087, respectively), as well as higher levels of body shame in both relapse groups compared to non-relapsed individuals (*p*=.091 and *p*=.074, respectively). Additionally, those in the Behavioral (Any Frequency) relapse group exhibited the following trends compared to non-relapsed individuals: a greater degree of difficulty stopping disordered eating thoughts (*p*=.098), lower levels of intuitive eating (*p*=.081), higher levels of depression (*p*=.079), higher levels of obsessionality (*p*=.070), and higher levels of perceived stress (*p*=.086). Of note, all patterns were in the expected direction for validation of the relapse constructs. Given the lack of statistical significance, however, further investigation is warranted into these trend-level relationships due to the small sample sizes and thus lower power and increased likelihood of type II errors.

We additionally inspected contingency tables, using Fisher's exact test due to the small sample, to examine whether individuals in the two relapse groups (Behavioral Relapse- Any

Frequency and DSM-5 Other Eating Disorders) were overrepresented within the "clinical" range on the EAT-26, CES-D, and DOCS compared to individuals who did not relapse. No individuals in the sample scored within the clinical range on the EAT-26, and therefore those analyses were not conducted. The Fisher's exact tests for the CES-D and the DOCS were not statistically significant (see Table 9), although the distribution was in the expected direction for both measures—individuals in the Behavioral Relapse (Any Frequency) condition more frequently scored in the clinical range (50%, or two of four) on the CES-D compared to those who did not relapse (19%, or four of 21), as well as on the DOCS (25%, or one of four) compared to those who did not relapse (0%, or zero of 21). Similarly, individuals in the DSM-5 Eating Disorders Relapse category more often scored in the clinical range (50%, or one of two) on the CES-D compared to those who did not relapse (19.1%, or four of 21) and more frequently scored in the clinical range on the DOCS (50%, or one of two) compared to those who did not relapse (0%, or zero of 21). Given the small cell sizes, however, caution must be taken when interpreting these patterns.

Discussion

This study sought to extend recent theoretical work exploring relapse in AN (Khalsa et al., 2017) by comparing several theory-driven definitions of relapse in a sample of individuals who have been deemed recovered from AN using a comprehensive definition of recovery (Bardone-Cone et al., 2010). This project is motivated by a desire to identify a standard definition of relapse for AN as a "benchmark" in the field to better understand this complex and chronic illness and to improve treatment outcomes. Additionally, establishing a standard definition of relapse can illuminate the developmental trajectories of symptom return following AN recovery, helping clinicians distinguish between those who are "slipping" back into AN from

those who are fully relapsed, and can inform timelines for intervention efforts in the highly sensitive period following AN treatment. Once consensus has been reached on a definition of relapse, researchers can use this construct to examine predictors of relapse in individuals with a history of AN to identify those who may need additional medical or psychological support when discharging from treatment or stepping down to a lower level of care.

Overall, only a handful of study participants (*n*=5) relapsed across the 1-year study period. While this is good news for our participants' emotional and physical health and provides further support for Bardone-Cone and colleagues' (2010) robust definition of recovery, we unfortunately were not able to test many of our relapse definitions for agreement with one another (Aim 2) or as grouping variables when comparing scores on clinically-significant measures between those who relapsed and those who did not (Aim 3). It is critical to note, however, that this study was greatly underpowered (*n*=26) to detect statistically significant and clinically meaningful differences between those who relapsed and those who did not relapse. As will be discussed later (see power analyses reported in Future Directions), the authors believe that these relapse definitions are strong and deserve to be examined in a larger sample in the future.

As previously noted, participants only met criteria for four out of eight relapse conditions over the study period: DSM-5 SubAN, DSM-5 Other Eating Disorders, Behavioral Relapse (Any Frequency), and Cognitive Relapse. Although we must be careful to draw conclusions from this, given our small sample, it is consistent with our Aim 1 hypothesis that the Behavioral Relapse (Any Frequency) and Cognitive Relapse definitions were most frequently endorsed by participants in the sample, with a total of five total participants meeting these definitions at either timepoint over the course of the study period. These two relapse conditions have low frequency

and low severity thresholds, therefore theoretically identifying those individuals who may be earlier in the relapse process. However, that two participants in the Behavioral Relapse (Any Frequency) group (IDs 144 and 450, see Table 4) later met relapse criteria for SubAN and OSFED-Purging Disorder is an interesting observation and suggests that perhaps individuals are more likely to experience eating disorder symptoms at full diagnostic threshold later (i.e., in this case, 12 months from their baseline visit), illustrating a developmental trajectory symptom return. In order to explore this further, future studies would need to standardize the time from which participants fully reached recovery to examine how long it takes for individuals to meet these lower severity definitions (e.g., Cognitive Relapse) and the higher severity definitions (e.g., DSM-5 SubAN).

That individuals in our sample did not meet criteria for many of the relapse definitions in the present study is, in the authors' opinion, not evidence to discard the constructs or the search for validating a definition of relapse in general. We believe that the Behavioral and Cognitive Relapse definitions are strong, theoretically driven, and represent a prevention-oriented approach to capturing the early stages of returning symptoms after an individual has reached recovery from AN. We also believe that Khalsa and colleagues' (2017) proposed relapse constructs are worth investigating further and comparing to other proposed definitions of relapse, especially to other relatively high severity relapse categories like DSM-5 AN. The lack of participants in our sample meeting criteria for Khalsa and colleagues' (2017) Full and Partial Relapse definitions is most likely attributable to their use of a high EDE-Q score (2 SDs of community norms) as well as the requirement of a BMI ≤ 18.5. In the current study, only one individual at T2 and two individuals at T3 had an EDE-Q score above 1 SD of age-matched norms. No individuals across the study period had an EDE-Q score above 2 SD of age-matched norms. Further, only one

individual had a BMI below 18.5 at any point in the study (this individual, ID 144, who had a BMI of 17.52, met criteria for DSM-5 SubAN definition at T3). It may be that these EDE-Q and BMI thresholds are too high, thereby only capturing those individuals who are already farther along in the relapse process and missing individuals who may be earlier in the relapse process and who could benefit from early intervention. However, larger samples are needed to determine if Khalsa and colleagues' (2017) proposed EDE-Q and BMI thresholds are too stringent or if they would be useful in identifying those who may be slipping from recovery toward the return of an eating disorder. In general, it is consistent with our Aim 1 hypotheses that the higher severity relapse definitions (e.g., DSM-5 AN, Khalsa Full Relapse) would be less frequently endorsed relative to the Behavioral and Cognitive Relapse categories.

Regarding our Aim 2 and Aim 3 hypotheses related to rates of agreement between relapse definitions and meaningfully distinguishing between relapsed and non-relapsed individuals using clinically-relevant self-report measures, little can be concluded from the present study given the small sample size. However, we did identify some statistically significant findings and have further made some observations of patterns that may be worth following up on in future studies. Independent samples t-tests revealed that individuals who met criteria for the Behavioral Relapse (Any Frequency) definition in our sample demonstrated statistically significantly higher rates of compelled or compulsive exercise and a greater number of hours of cardio and strength training compared to those who did not relapse, and they were more likely cite stress management or socializing as their motivations for engaging in exercise. Individuals who met criteria for DSM-5 Eating Disorders (including SubAN and OSFED-Purging Disorder) reported higher levels of depression compared to those who did not relapse and, similar to the Behavioral Relapse (Any Frequency) group, also reported a greater number of hours of exercise compared to those who

did not relapse. With regard to constructs with cutoffs denoting clinically concerning scores, although none of the inferential tests were statistically significant, individuals in both relapse categories more frequently scored in the clinical range on the CES-D and the DOCS. These sets of findings, significant and trend-level, were in the expected direction (i.e., greater pathology in the relapse group than in the non-relapsed group) and warrant future exploration with larger samples.

Case Study

Our study's small sample size, although a limitation, also allows for exploration of relapse at the individual level, and case examples may provide insight into how an individual's symptoms unfold over time. One of the individuals who relapsed at both T2 and T3 (ID 450, see Table 4) was a 43-year-old heterosexual divorced white female whose lifetime lowest BMI (per her report) was 12.84. Her BMI at baseline was 20.41. At T2, she reported having fasted once over the past month but denied purging behaviors. Her T2 BMI was in the normal range (20.99) and her EDE-Q score (2.54) was not outside of the 1 SD range (2.69) or the 2 SD range (3.71) for her age. Given these factors, she met relapse criteria for the Behavioral Relapse (Any Frequency) definition at T2. In addition to fasting, at T2 she reported engaging in excessive exercise multiple times per week over the past few months, though her exercise behaviors were not captured in the definition of Behavioral Relapse. Approximately six months later at the T3 visit, this same individual reported 14 instances of self-induced vomiting in the three months prior to assessment (in addition to excessive exercise) and met diagnostic criteria for Purging Disorder (an OSFED diagnosis). Her BMI was in the normal range (21.50) and her EDE-Q score (2.79) was outside the 1 SD limit for her age range (2.69). Because she was engaging in disordered eating behaviors *and* endorsed clinically significant disordered eating disorder

thoughts on the EDE-Q, she did not meet for either the Behavioral Relapse or Cognitive Relapse conditions. Further, because her EDE-Q score was not outside the 2 SD limit for her age range, she did not meet criteria for either Khalsa Partial or Full Relapse.

This case example illustrates the complexity of the relationships among physical, behavioral, and cognitive symptom domains of AN, the way these domains interact over time, and the frequency of disordered eating behaviors and cognitions in individuals with a history of AN. For example, the choice to exclude excessive exercise from the weight loss behaviors examined in this sample was informed by the behaviors included in Bardone-Cone and colleagues' (2010) definition of recovery—the authors noted that due to the nebulous definition of what qualifies as "excessive" physical activity, which ultimately introduces error from individual clinical judgment, there were concerns that it would not be a reliable variable to include for an indicator of recovery (A. Bardone-Cone, personal communication, April 20, 2021). However, in the example of participant 450 described above, this individual's frequency of excessive exercise would have resulted in their categorization in the Behavioral Relapse (1x/Week Frequency) group if that relapse construct included excessive exercise in addition to other disordered eating behaviors (i.e., fasting, binge eating, purging). Further, the participant's engagement in high frequency excessive exercise at T2 may have "flagged" her for a higher level of care before she began purging routinely at T3. This may be evidence in support of evaluating an individuals' excessive exercise habits in future work exploring relapse, so long as researchers clearly define what is meant by "excessive." This case example also illustrates the importance of monitoring symptom change over time—in the sensitive period following treatment, clinicians should continuously assess for other disordered eating and weight loss behaviors (e.g., purging), even in individuals who were not recently engaging in these behaviors. These are only a few of

the challenges we face as researchers and clinicians in validating a definition of relapse, though further exploration is warranted for clarity in research projects and improving clinical care and patient outcomes.

Strengths and Limitations

This study's longitudinal design, use of gold-standard measures (e.g., EDE-Q), and advancement of prior work related to eating disorder recovery (Bardone-Cone et al., 2010) and relapse (Khalsa et al., 2017) are some of its strengths. The use of a comprehensive definition of recovery (per Bardone-Cone et al., 2010) is also a strength, as is the use of quality of life and psychosocial functioning variables (in addition to psychopathology variables) as holistic health domains for use in the aim of validating relapse constructs. Perhaps most importantly, the relapse constructs proposed in the present study are theoretically strong, contribute to existing discussions about the illness course of AN, and offer directions for future investigation. Work building on these theoretical constructs may ultimately have significant implications for clinical practice across treatment settings (e.g., primary care, outpatient psychotherapy, inpatient treatment). For example, if a relapse definition with a more "subtle" and lower severity threshold (e.g., Cognitive Relapse) relative to higher-severity constructs (e.g., DSM-5 AN) emerges as a valid indicator of relapse, this definition could be disseminated to care settings in which providers may have less training in identifying eating disorder symptomatology but a high likelihood of encountering individuals with eating pathology, and which provide routine points of contact. Primary care providers may then be encouraged to administer the EDE-Q at regular outpatient visits to capture the earliest return of symptoms in individuals with a history of AN who may not be in contact with a therapist or dietician, and then get these individuals connected with more specialized care. Similarly, if the constructs with lower symptom severity thresholds

(e.g., Behavioral Relapse, Cognitive Relapse) are strong, therapists treating individuals in recovery from AN may pay closer attention to the frequency of weight loss behaviors or eating disorder cognitions at routine outpatient visits to monitor a client's "lapse" or "relapse." To this end, it is the authors' hope that future researchers continue to explore these ideas.

The primary limitation of the current study is its small sample size. We anticipated a small *N*, considering the low base rate of AN and the relatively low rates of full recovery. Correspondingly, we ran an a priori power analysis for independent sample *t*-tests (for Aim 3) using G*Power (Faul et al., 2007; Faul et al., 2009). Based on an estimate for the number of individuals with a lifetime history of AN who provided T1 and T2 data (n=168) and based on prior research using this recovery definition that resulted in 21% of a sample with an eating disorder history meeting criteria for full recovery (Bardone-Cone et al., 2010), we anticipated about 35 participants to be fully recovered at T1. Using G*Power with a sample size of 35, alpha of .05, and a medium effect size of 0.3, power was expected to be 0.57. for a one-tailed *t*-test. For a large effect size of 0.5, power was expected to be 0.96. However, our actual sample size (N = 26) was lower than our anticipated sample size of 35—we believe this is because the study on which this estimation was based (Bardone-Cone et al., 2010) recruited patients from an outpatient primary care clinic, which likely treated, on average, lower-severity patients than those seen in the present study, who were recruited from inpatient care. Therefore, patients in the original study were more likely to have achieved full recovery than those in the present study, thereby providing a larger baseline sample.

Based on the number of individuals in the current study who met relapse definitions and those who did not (approximately a 1:10 ratio of relapsed:not relapsed), we ran another power analysis using G*Power (Faul et al., 2007) to determine adequate group sizes. This analysis

revealed that to detect large-sized effects (i.e., 0.5), approximately 302 individuals would be needed to conduct independent samples pairwise comparisons, with 27 individuals in the relapsed condition and 275 in the not-relapsed condition (calculation based on p=0.05, power=0.80). To detect medium-sized effects (i.e., 0.3), approximately 834 individuals would be needed, with 76 in the relapsed condition and 758 in the not-relapsed condition. As groups of this size are unattainable for a single-site study, future efforts to examine relapse in a sample of individuals in full recovery from AN should employ multi-site data collection of those successfully treated in and discharged from inpatient weight restoration units to ensure sufficient group sizes for analysis.

The present study is also limited by its methodology, as the original study was not designed to answer questions related to these specific relapse constructs. As such, the time frames of the measures used to define some relapse constructs did not align in every case (e.g., EDE-Q assesses cognitions over the past 28 days while the Khalsa full relapse definition sets a duration period of three months). Future work exploring these relapse constructs should mind the assessment time periods for all measures and research tools used to ensure the proposed relapse period (e.g., 3 months) is fully assessed—this will paint a more detailed picture of how eating disorder symptoms unfold over time and may reveal important temporal relationships between the frequencies of behaviors and cognitions. Additionally, participants in the original study were included in the sample if they were "at least" one year out of intensive treatment, though there was variation in their time from their most recent hospitalization. Future work should also aim to standardize the T1 data collection in relation to discharge from treatment—for example, capturing only those who discharged from an inpatient care facility in the past year to begin assessing relapse in the sensitive post-treatment window (Berends et al., 2016), and additionally

assess the time at which individuals met criteria for recovery (per Bardone-Cone et al.'s [2010] definition). Doing so may capture more individuals who meet definitions of relapse.

Finally, our sample was overwhelmingly homogenous with regard to participant identities (e.g., gender, race, ethnicity, etc.). In order to fully understand the course of AN and how it manifests across individuals, future work should intentionally recruit a more diverse sample, making special efforts to including non-white, non-binary, and differently-abled bodies, and further explore how lived experiences of identity-related stress or discrimination may affect rates of relapse or relapse phenomenology.

Future Directions

Results from the present study suggest many future research directions that can further build on this foundation of knowledge. Ideally, a study seeking to validate a definition of relapse should examine symptoms longitudinally over the course of two years or more following recovery, as this has been identified as a meaningful window of risk after successful treatment (Berends et al., 2018), with more frequent assessments on a monthly, weekly, or daily basis (e.g., using ecological momentary assessment) to capture the granularity of the relapse "process" unfolding over time. This methodology would allow future researchers to examine relapse constructs more dimensionally instead of categorically—for example, assessing weight loss/gain trajectories, frequency of behaviors, changes in EDE scores, and comorbid psychopathology symptoms—to measure *degree* of relapse with greater sensitivity (e.g., yielding a numerical score on a multidimensional relapse scale with thresholds for "mild," "moderate," or "severe" relapse). This approach may also reveal temporal relationships between eating disorder symptomatology and other variables (e.g., depression and anxiety symptoms, stress, major life events, etc.), addressing important questions about how frequently eating disorder relapse leads

to other illness or life stress, how frequently illness or life stress triggers eating disorder relapse, and how frequently these phenomena occur concurrently. As previously noted, it will be necessary to conduct these studies across multiple sites to ensure adequate participant recruitment for analyses.

With a more granular longitudinal methodology as described above, researchers could employ a Kaplan-Meier "survival" curve to describe *time-to-event* data using time as a continuous variable and different relapse definitions as categorical variables (Ranstam & Cook, 2017), illustrating the proportion of individuals who are at risk for relapse over a specific period of time (e.g., two years). This approach could be combined with Cox proportional-hazards models to identify treatment-related predictors of different survival curves. Analysis using this approach could examine duration of hospital stay or timing of follow-up visits after discharge from an inpatient unit as possible predictors of trajectories for individuals with a history of AN.

Arguably, an important reason for clarifying relapse definitions is to then be able to identify predictors of these states of relapse, not just from treatment experiences, but from a wide array of potential predictors. For example, what characteristics (related to identity variables, comorbid psychopathology, personality/temperament, social support, relationship stress, adverse life experiences, resilience factors, etc.) predict different profiles of relapse after achieving recovery from AN? What characteristics predict relapse to another non-AN eating disorder in those with a history of AN? How does age, age of illness onset, or duration of illness affect likelihood of relapse? How long do relapse "episodes" last in the absence of treatment, on average, and what factors affect the length or course of these "episodes"? The answers to these questions can inform AN intervention and relapse prevention strategies by helping providers identify those who may be at increased risk for relapse following discharge from the hospital.

Additionally, further exploration of relapse in AN is necessary to improve both treatment outcomes and our understanding of the illness. Over the past several years, researchers have begun to use the qualifier "severe and enduring" to refer to AN that is persistent (e.g., lasting at least seven years) and resistant to treatment (Broomfield et al., 2017), and it is believed that up to 30% of individuals with AN demonstrate this illness pattern (Wonderlich et al., 2020). However, establishing a standardized definition of relapse is necessary to clarify the boundaries of this illness qualifier—for example, is "severe and enduring" a qualitatively unique presentation of AN, distinct from the illness pattern of individuals oscillating episodically between discrete states of recovery and relapse?

Finally, it is possible that, as researchers and clinicians, we are overlooking factors that could be contributing to the "enigmatic persistence" of AN (Walsh, 2013). Qualitative methodology (e.g., interviewing patients about their subjective experience of illness, recovery, and relapse) could reveal unexpected themes and provide novel directions for future work examining the developmental trajectories and treatment of AN.

Conclusion

In sum, the present study builds on existing theoretical work on AN relapse (Khalsa et al., 2017) and recovery (Bardone-Cone et al., 2010) in an effect to establish a definition of relapse with both empirical and clinical utility. Although this study's small sample size limits the conclusions we can draw from the findings, the authors believe these proposed relapse constructs are worth exploring further through collaborative, multi-site work. Further examination will help identify an empirically useful "benchmark" for relapse in research studies of AN, clarify developmental trajectories and stages of illness in AN, and reveal AN symptom (e.g., cognitive, behavioral, physical) combinations and levels of severity that meaningfully distinguish

individuals on clinically-relevant variables (e.g., depressed mood, intuitive eating) to identify those who, following AN recovery, may need a higher level of care. It is the authors' hope that others will continue this important work.

APPENDIX 1: DSM-5 EATING DISORDERS DIAGNOSTIC CRITERIA

DSM-5 Eating Disorders Diagnostic Criteria (American Psychiatric Association, 2013)

Anorexia Nervosa

A) Restriction of energy intake relative to requirements, leading to a significantly low body weight in the context of age, sex, developmental trajectory, and physical health. *Significantly low weight* is defined as a weight that is less than minimally normal or, for children and adolescents, less than that minimally expected.

B) Intense fear of gaining weight or of becoming fat, or persistent behavior that interferes with weight gain, even though at a significantly low weight.

C) Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or persistent lack of recognition of the seriousness of the current low body weight.

Atypical Anorexia Nervosa (Other Specified Feeding or Eating Disorder)

A) All of the above criteria for anorexia nervosa are met, except that despite significant weight loss, the individual's weight is within or above the normal range.

Bulimia Nervosa

A) Recurrent episodes of binge eating characterized by both of the following:

1. Eating, within a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than what most people would eat during a similar period of time and under similar circumstances.
2. A sense of lack of control over the eating episode (e.g., a feeling that one cannot stop eating or control how much one is eating).

B) Recurrent inappropriate compensatory behavior to prevent weight gain, such as self-induced vomiting, misuse of laxatives, diuretics, or other medications, fasting, or excessive exercise.

C) The binge eating and inappropriate compensatory behaviors both occur, on average, at least once a week for three months.

D) Self-evaluation is unduly influenced by body shape and weight.

E) The disturbance does not occur exclusively during episodes of anorexia nervosa.

Binge-Eating Disorder

A) Recurrent episodes of binge eating characterized by both of the following:

1. Eating, within a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than what most people would eat during a similar period of time and under similar circumstances.
2. A sense of lack of control over the eating episode (e.g., a feeling that one cannot stop eating or control how much one is eating).

B) Binge-eating episodes are associated with three (or more) of the following:

1. Eating much more rapidly than normal.
2. Eating until feeling uncomfortably full.
3. Eating large amounts of food when not feeling physically hungry.
4. Eating alone because of feeling embarrassed by how much one is eating.
5. Feeling disgusted with oneself, depressed or very guilty afterward.

C) Marked distress regarding binge eating is present.

D) Binge eating occurs, on average, at least once a week for three months.

E) Binge eating is not associated with the recurrent use of inappropriate compensatory behaviors as in bulimia nervosa and does not occur exclusively during the course of bulimia nervosa or anorexia nervosa

Other Specified Feeding and Eating Disorders

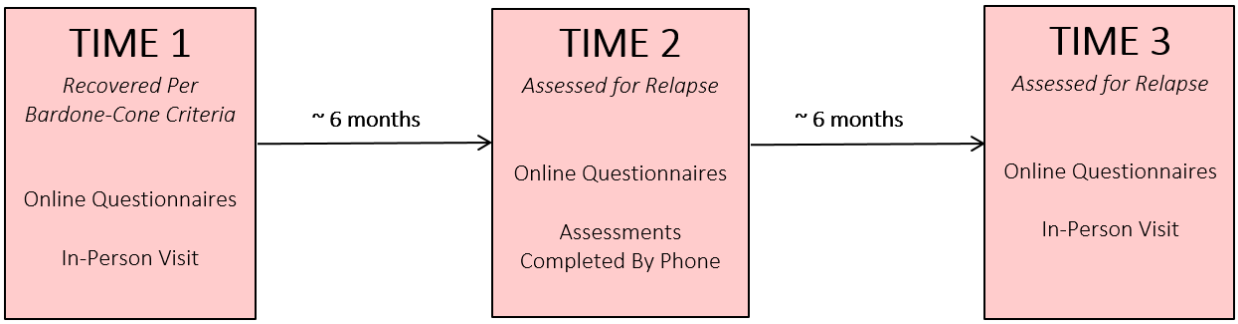
A) Feeding or eating behaviors that cause clinically significant distress and impairment in social, occupational, or other important areas of functioning, but do not meet the full criteria for any of the other feeding and eating disorders.

Examples of presentations:

1. Bulimia nervosa (of low frequency and/or limited duration): All of the criteria for bulimia nervosa are met, except that the binge eating and inappropriate compensatory behaviors occur, on average, less than once a week and/or for less than 3 months.
2. Binge-eating disorder (of low frequency and/or limited duration): All of the criteria for binge-eating disorder are met, except that the binge eating occurs, on average, less than once a week and/or for less than 3 months.
3. Purging disorder: Recurrent purging behavior to influence weight or shape (e.g., self-induced vomiting; misuse of laxatives, diuretics, or other medications) in the absence of binge eating.

APPENDIX 2: FIGURE 1- PROCEDURE TIMELINE

Figure 1
Procedure Timeline



APPENDIX 3: TABLES

Table 1

Relapse Constructs, Operationalization of Symptoms, and Assessment Tools Used to Assess Symptoms

RELAPSE DEFINITION			OPERATIONALIZATION	SYMPTOM DURATION	ASSESSMENT TOOLS
DSM-5 Diagnostic Criteria Relapse Constructs	Anorexia Nervosa and Subthreshold Anorexia Nervosa	DSM-5 Diagnostic Criteria for Anorexia Nervosa	Meets DSM-5 diagnostic criteria[1]	1 month[2]	SCID
		DSM-5 Diagnostic Criteria for Subthreshold / "Atypical" Anorexia Nervosa	Meets DSM-5 diagnostic criteria[1]	1 month[2]	
	Non-AN Eating Disorders	DSM-5 Criteria for Bulimia Nervosa, Binge Eating Disorder, or Other Specified Feeding and Eating Disorder[3]	Meets DSM-5 diagnostic criteria[1]	3 months[2]	

RELAPSE DEFINITION			OPERATIONALIZATION			SYMPTOM DURATION	ASSESSMENT TOOLS
			WEIGHT STATUS	BEHAVIORS	COGNITIVE		
Other Theoretical Relapse Constructs	Khalsa Definitions	Khalsa Full Relapse	BMI ≤ 18.5	Significant (on average, 1x/week for past 3 months) restricting[6], bingeing, or purging present	Significant fear of gaining weight or disturbance in body image (EDE-Q[5] global score ≥ 2 SD of normal)	3 months[4]	BMI calculated using height/weight measured using digital scale and stadiometer (at T1 and T3) or self-reported height/weight (at T2). Behavioral frequency of fasting, bingeing, or purging (e.g., vomiting or laxative or diuretic use) was assessed using annotated calendars administered by study staff. Fear of weight gain and body image disturbance assessed using EDE-Q.
		Khalsa Partial Relapse	BMI ≤ 18.5	Restricting[6], bingeing, or purging present (<u>any</u> frequency of behaviors in past month)	Significant fear of gaining weight or disturbance in body image (EDE-Q[5] global score ≥ AA2 SD of normal)	1 month	
	Walsh & Bardone-Cone Definitions	Cognitive Relapse	BMI ≥ 18.5	No bingeing, purging, *or* fasting in the past 3 months	EDE-Q global score > 1 SD of age-matched community norms	3 months[4]	

		Behavioral Relapse	BMI ≥ 18.5	Bingeing, purging, or fasting, on average, 1x/week for 3 months	No clinically significant body image disturbance or fear of weight gain (EDE-Q global score ≤ 1 SD of age-matched community norms)	3 months[4]	
				Bingeing, purging, or fasting (any frequency of behaviors) in past 3 months			

[1]See Appendix for DSM-5 diagnostic criteria.

[2]The SCID assesses symptoms over variable timeframes for each diagnosis. Anorexia nervosa and subthreshold anorexia nervosa were assessed during the past month, bulimia nervosa and binge-eating disorder over the past 3 months, and other specified feeding and eating disorders (with the exception of subthreshold anorexia nervosa) over the past 3 months.

[3]Other specified feeding and eating disorders (OSFED) includes OSFED-subthreshold BN, OSFED-subthreshold BED, and OSFED-purging disorder.

[4]The Eating Disorder Examination Questionnaire (EDE-Q) assesses symptoms within the past 28 days, though behavioral frequency was assessed over the past 3 months. For Khalsa and colleagues' (2017) full relapse definition and the cognitive relapse and behavioral relapse definitions, we were able to assess the behavioral component over the past 3 months but only assess the cognitive component over the past month.

[5]Khalsa and colleagues suggested using the EDE (Eating Disorder Examination), though in the present study we used the EDE-Q.

[6] Study staff inquired about fasting, defined as intentionally (and not for religious reasons) going without eating for a 24-hour period, which in the current study was used as a proxy for "restricting."

Table 2
Measures Used to Assess Validity of Relapse Definitions and Internal Consistency Data in the Study Sample

CONSTRUCT	MEASURES	COEFFICIENT ALPHA T2	COEFFICIENT ALPHA T3
Disordered Eating Attitudes and Behaviors	Eating Attitudes Test-26 (EAT-26)	.76	.75
	Obsessionality related to weight/shape/eating (amount of time spent thinking about thoughts)	.76	.82
	Obsessionality related weight/shape/eating (difficulty stopping thoughts)	.85	.84
	Body Shame Subscale of Objectified Body Consciousness Scale (OBCS)	.78	.77
	Body Surveillance Subscale of OBCS	.87	.82
	Thinness and Restricting Expectancy Inventory (brief version) (TREI)	.94	.89
	Reasons for Exercise Inventory (REI): Fitness and Health	.95	.89
	Reasons for Exercise Inventory (REI): Appearance and Weight	.86	.81
	Reasons for Exercise Inventory (REI): Stress and Mood Management	.82	.83
	Reasons for Exercise Inventory (REI): Socializing	.91	.88
	Compulsive Exercise Questions from Delphi Panel	.95	.92
Positive Eating- and Body-Related Attitudes	Body Appreciation Scale (BAS)	.92	.92
	Intuitive Eating Scale (IES)	.92	.91
Non-Eating Disorder Psychopathology	Center for Epidemiological Studies Depression Scale (CES-D)	.81	.90
	Trait Anxiety Scale of Spielberger State-Trait Anxiety Inventory (STAI-T)	.89	.90
	Dimensional Obsessive-Compulsive Scale (DOCS)	.90	.90

Personality Traits	Distress Tolerance Scale (DTS)	.94	.93
	Negative Urgency Subscale of UPPS Impulsive Behavior Scale (UPPS)	.80	.84
Psychosocial Functioning and Quality of Life	Social Health Subscale of the Patient-Report Outcomes Measurement Information System (PROMIS)	.96	.96
	Perceived Stress Scale (PSS)	.84	.85

Table 3
Bardone-Cone et al. (2010)[3] Eating Disorder Recovery Definition and Assessment Tools

RECOVERY COMPONENT	OPERATIONALIZATION	ASSESSMENT TOOL
Diagnosis	No current eating disorder diagnosis (AN, BN, BED, OSFED[1])	SCID
Physical Recovery	BMI $\geq$ 18.5 kg/m^2	Measured weight and height using digital scale and stadiometer[1]
Behavioral Recovery	No binge eating, vomiting, laxatives, or fasting within past 3 months	Annotated calendars administered by study staff assessing frequency of behaviors over past 3 months
Cognitive Recovery	All four EDE-Q subscales within 1 SD of age-matched community norms[2]	EDE-Q

Note. AN=anorexia nervosa, BN=bulimia nervosa, BED=binge-eating disorder, OSFED=other specified feeding and eating disorder, BMI=Body Mass Index, SCID=Structured Clinical Interview for DSM, EDE-Q=Eating Disorders Examination-Questionnaire.

[1] Participant weight and height were measured for Time 1 and Time 3 during participants' study visits. At Time 2, which participants completed remotely on the computer and on phone interviews, height and weight were self-reported.

[2]Norms based on Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy*, *44*, 53-62. https://doi.org/10.1016/j.brat.2004.12.003.

[3]Bardone-Cone, A.M., Harney, M.B., Maldonado, C.R., Lawson, M.A., Robinson, D.P., Smith, R., & Tosh, A. (2010). Defining recovery from an eating disorder: Conceptualization, validation, and examination of psychosocial functioning and psychiatric comorbidity. *Behaviour Research and Therapy*, *48*, 194-202. doi: 10.1016/j.brat.2009.11.001

Table 4

Relapse Definitions Met by Individual Participants Across Both Study Timepoints

ID	DSM-5 AN		DSM-5 SubAN		DSM-5 Other Eating Disorders		Khalsa Full		Khalsa Partial		Behavioral (1x/Week)		Behavioral (Any Frequency)		Cognitive	
	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3
110																
112																
126																
132														✓		
135																
144				✓									✓			
154																
165																
167													✓			
183																
187																
192																
196																
278																
302																
331																
341																
344																
369																
371																
375															✓	
404																
417																
423																
450						✓							✓			
488																

Note. The rows of individual who did not relapse at any point during the study period are shaded. The rows of individuals who met criteria for a relapse definition at either T2, T3, or at both timepoints are unshaded. The relapse definition for which they met criteria is indicated with a checkmark (✓).

Table 5
Relapse Definitions Met at Time 2 and Time 3, Reported Separately

Relapse Definitions Met at Time 2 (*n*=4)

Relapse Definition	*n*	% of total sample (*N*=26)
DSM-5 AN	0	0%
DSM-5 SubAN	0	0%
DSM-5 Other EDs	0	0%
Khalsa Full Relapse	0	0%
Khalsa Partial Relapse	0	0%
Behavioral Relapse (1x/Week)	0	0%
Behavioral Relapse (Any Frequency)	3	11.5%
Cognitive Relapse	1	3.8%

Note. AN=anorexia nervosa, EDs=eating disorders.

Relapse Definitions Met at Time 3 (*n*=3)

Relapse Definition	*n*	% of total sample (*N*=26)
DSM-5 AN	0	0%
DSM-5 SubAN	1	3.8%
DSM-5 Other EDs	1	3.8%
Khalsa Full Relapse	0	0%
Khalsa Partial Relapse	0	0%
Behavioral Relapse (1x/Week)	0	0%
Behavioral Relapse (Any Frequency)	1	3.8%
Cognitive Relapse	0	0%

Note. AN=anorexia nervosa, EDs=eating disorders.

Table 6

Cohen's Kappa for Relapse Definitions at T2

T2		DSM-5 AN	DSM-5 SubAN	DSM-5 Other EDs	Khalsa Full	Khalsa Partial	Behavioral (1x/wk)	Behavioral (Any Freq)	Cognitive
		N=0	N=0	N=0	N=0	N=0	N=0	N=3	N=1
DSM-5 AN	N=0								
DSM-5 SubAN	N=0	--							
DSM-5 Other EDs	N=0	--	--						
Khalsa Full	N=0	--	--	--					
Khalsa Partial	N=0	--	--	--	--				
Behavioral (1x/wk)	N=0	--	--	--	--	--			
Behavioral (any)	N=3	--	--	--	--	--	--		
Cognitive	N=1	--	--	--	--	--	--	Y=0, N=22 Kappa: -.061	

Note. For comparisons of rates between two relapse conditions, "Y" refers to the number of participants who met *both* relapse conditions and "N" refers to the number of participants who met *neither* relapse condition.
AN=anorexia nervosa, EDs=eating disorders.

Table 7

Cohen's Kappa for Relapse Definitions at T3

T3		DSM-5 AN	DSM-5 SubAN	DSM-5 Other EDs	Khalsa Full	Khalsa Partial	Behavioral (1x/wk)	Behavioral (any)	Cognitive
		N=0	N=1	N=1	N=0	N=0	N=0	N=1	N=0
DSM-5 AN	N=0								
DSM-5 SubAN	N=1	--							
DSM-5 Other EDs	N=1	--	Y=0, N=24 Kappa = -.040						
Khalsa Full	N=0	--	--	--					
Khalsa Partial	N=0	--	--	--	--				
Behavioral (1x/wk)	N=0	--	--	--	--	--			
Behavioral (any)	N=1	--	Y=0, N=24 Kappa = -.040	Y=0, N=24 Kappa = -.040	--	--	--		
Cognitive	N=0	--	--	--	--	--	--	--	

Note. For comparisons of rates between two relapse conditions, "Y" refers to the number of participants who met *both* relapse conditions and "N" refers to the number of participants who met *neither* relapse condition.
AN=anorexia nervosa, EDs=eating disorders.

Table 8

Independent Samples t-tests Between Individuals Who Met Relapse Definitions and Those Who Did Not Relapse Over the Study Period

	Behavioral Relapse (Any Frequency)			DSM-5 Eating Disorder Diagnoses (SubAN and Purging Disorder)		
Measure	**Never Relapsed (n=21)**	**Relapsed (n= 4)**	**Sig**	**Never Relapsed (n=21)**	**Relapsed (n=2)**	**Sig**
EAT-26	4.86 (4.35)	9.25 (5.31)	*p*=.086^	4.86 (4.35)	11.00 (8.48)	*p*=.087^
Obsessionality Related to Weight/Shape/Eating (Amt of Time Spent Thinking About)	1.71 (0.59)	1.87 (0.59)	*p*=.612	1.71 (0.59)	2.25 (0.71)	*p*=.235
Obsessionality Related to Weight/Shape/Eating (Difficulty Stopping Thoughts)	1.86 (0.66)	2.50 (0.79)	*p*=.098^	1.86 (0.66)	2.37 (0.53)	*p*=.302
OBCS: Shame	3.15 (0.92)	4.06 (1.05)	*p*=.091^	3.15 (0.92)	4.50 (1.59)	*p*=.074^
OBCS: Surveillance	4.23 (0.98)	4.12 (1.07)	*p*=.840	4.23 (0.98)	4.19 (1.86)	*p*=.951
Thinness and Restricting Expectancy Inventory	28.26 (10.38)	29.75 (13.22)	*p*=.803	28.26 (10.38)	26.00 (16.97)	*p*=.780
Reasons for Exercise: Fitness/Health	45.52 (14.58)	54.00 (1.63)	*p*=.266	45.52 (14.58)	55.00 (1.41)	*p*=.379
Reasons for Exercise: Appearance/Weight	34.52 (11.36)	42.25 (12.20)	*p*=.232	34.52 (11.36)	44.50 (16.26)	*p*=.262
Reasons for Exercise: Stress Management	19.80 (5.47)[1]	24.25 (0.96)[1]	***p*=.003****	19.80 (5.47)	24.00 (1.41)	*p*=.302
Reasons for Exercise: Socializing	6.08 (2.96)	9.75 (1.26)	***p*=.026***	6.08 (2.96)	9.00 (1.41)	*p*=.191
Compulsive Exercise: Sum of Hours Cardio & Strength	5.01 (3.17)	10.00 (3.74)	***p*=.010***	5.01 (3.17)	12.00 (2.83)	***p*=.007****
Compulsive Exercise: Thoughts About Exercise	38.39 (23.48)[1]	73.59 (9.38)[1]	***p*<.001*****	38.39 (23.48)	73.25 (8.13)	*p*=.053^
Body Appreciation Scale	3.77 (0.60)	3.67 (0.51)	*p*=.755	3.77 (0.60)	3.42 (0.49)	*p*=.435
Intuitive Eating Scale	77.62 (12.24)	65.50 (11.70)	*p*=.081^	77.62 (12.24)	68.00 (9.90)	*p*=.296
Depression (CES-D)	10.07 (5.75)	16.00 (6.31)	*p*=.079^	10.07 (5.75)	19.00 (5.66)	***p*=.048***
Trait Anxiety (STAI-T)	42.86 (7.96)	45.25 (9.94)	*p*=.600	42.86 (7.96)	43.50 (9.19)	*p*=.915
Dimensional Obsessive-Compulsive Scale	5.19 (4.67)	10.25 (6.13)	*p*=.070^	5.19 (4.67)	11.00 (9.90)	*p*=.135
Distress Tolerance Scale	2.71 (0.79)	3.22 (0.54)	*p*=.234	2.71 (0.79)	3.10 (0.21)	*p*=.498
Negative Urgency (UPPS)	2.08 (0.41)	2.42 (0.30)	*p*=.137	2.08 (0.41)	2.50 (0.47)	*p*=.188
Perceived Stress scale	21.74 (5.36)	27.50 (8.54)	*p*=.086^	21.74 (5.36)	24.00 (12.73)	*p*=.611

PROMIS Social Support	64.86 (10.87)	57.00 (12.88)	*p*=.210	64.86 (10.87)	49.00 (0.00)	*p*=.056^

Note. ^*p*<.10, **p*<.05, ***p*<.01, ****p*<.001.

AN=anorexia nervosa, EAT-26=Eating Attitudes Test- 26 Item, OBCS=Objectified Body Consciousness Scale, CES-D=Center for Epidemiological Studies-Depression Scale, STAI-T=State-Trait Anxiety Inventory-Trait Measure.

The Reasons for Exercise Inventory was optional at both Time 2 and Time 3 ("If you never exercise, please skip this section"). At both Time 2 and 3, two participants did not complete the measure. These participants were both in the Never Relapsed category, so the total number of participants in the Never Relapsed category who completed the REI at Time 2 and Time 3 was *n*=19.

[1]According to Levene's Test for Equality of Variances, equal variances were not assumed and the t-statistic reported in the text is the one that that accompanies a significant Levene's test. In all other cases, t-statistics reported in the text are based on assumed equal variances (i.e., non-significant Levene's test).

Table 9

Contingency Tables for Measures for Which There Are Clinically Meaningful Cutoff Scores

	Behavioral Relapse (Any Frequency)			DSM-5 Eating Disorder Diagnoses (SubAN and Purging Disorder)		
	Not Relapsed (n=21)	**Relapsed (n=4)**	**Sig (Fisher's Exact)**	**Not Relapsed (n=21)**	**Relapsed (n=2)**	**Sig (Fisher's Exact)**
EAT-26 Subclinical Range (LT 20)	100% (21)	100% (4)	--	100% (21)	100% (2)	--
EAT-26 Clinical Range (GE 20)	0% (0)	0% (0)		0% (0)	0% (0)	
CES-D Subclinical Range (LT 16)	81.0% (17)	50.0% (2)	*p* = .234 (n.s.)	80.9% (17)	50% (1)	*p* = .395 (n.s.)
CES-D Clinical Range (GE 16)	19.0% (4)	50.0% (2)		19.1% (4)	50% (1)	
DOCS Subclinical Range (LT 18)	100% (21)	75% (3)	*p* = .160 (n.s.)	100% (21)	50% (1)	*p* = .087 (n.s.)
DOCS Clinical Range (GE 18)	0% (0)	25% (1)		0% (0)	50% (1)	

Note. AN=anorexia nervosa, EAT-26=Eating Attitudes Test-26 Item, CES-D=Center for Epidemiological Studies-Depression Scale, DOCS=Dimensional Obsessive-Compulsive Scale, LT=less than, GE=greater than or equal to.

REFERENCES

Aardoom, J.J., Dingemans, A.E., Slof Op't Landt, M.C.T., & Van Furth, E.F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors, 13,* 305-309. https://doi.org/10.1016/j.eatbeh.2012.09.002

Abramowitz, J.S., Deacon, B.J., Olatunji, B.O., Wheaton, M.G., Berman, N.C., Losardo, D., Timpano, K.R., McGrath, P.B., Riemann, B.C., Adams, T., Bjorgvinsson, T., Storch, E.A., & Hale, L.R. (2010). Assessment of obsessive-compulsive symptom dimensions: Development and evaluation of the Dimensional Obsessive-Compulsive Scale. *Psychological Assessment, 22*, 180-198. DOI: 10.1037/a0018260

Ackard, D.M., Richter, S.A., Egan, A.M., & Cronemeyer, C.L. (2014). What does remission tell us about women with eating disorders? Investigating applications of various remission definitions and their associations with quality of life. *Journal of Psychosomatic Research,* 76, 12-18. https://doi.org/10.1016/j.jpsychores.2013.10.002

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anestis, M.D., Selby, E.A., & Joiner, T.E. (2007). The role of urgency in maladaptive behaviors. *Behavior Research and Therapy, 45*, 3018-3029. DOI: 10.1016/j.brat.2007.08.012

Arcelus, J., Mitchell, A.J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. *Archives of General Psychiatry, 68*, 724-731. doi: 10.1001/archgenpsychiatry.2011.74

Avalos, L.C. & Tylka, T.L. (2006). Exploring a model of intuitive eating with college women. *Journal of Counseling Psychology, 53*, 486-497. DOI: 10.1037/0022-0167.53.4.486

Avalos, L., Tylka, T.L., & Wood-Barcalow (2005). The Body Appreciation Scale: Development and psychometric evaluation. *Body Image, 2*, 285-297. doi:10.1016/j.bodyim.2005.06.002

Bardone-Cone, A. M., Alvarez, A., Gorlick, J., Koller, K. A., Thompson, K., & Miller, A. (2019). Longitudinal follow-up of a comprehensive operationalization of eating disorder recovery: Concurrent and predictive validity. *International Journal of Eating Disorders, 52,* 1052-1057. doi.org/10.1002/eat.23128

Bardone-Cone, A.M., Butler, R.M., Balk, M.R., & Koller, K.A. (2016). Dimensions of impulsivity in relation to eating disorder recovery. *International Journal of Eating Disorders, 49*, 1027-2031. doi:10.1002/eat.22579.

Bardone-Cone, A.M., Harney, M.B., Maldonado, C.R., Lawson, M.A., Robinson, D.P., Smith, R., & Tosh, A. (2010). Defining recovery from an eating disorder: Conceptualization, validation, and examination of psychosocial functioning and psychiatric comorbidity. *Behaviour Research and Therapy, 48*, 194-202. doi: 10.1016/j.brat.2009.11.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. (Methodological), 57,* 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berends, T., Boonstra, N., & van Elburg, A. (2018). Relapse in anorexia nervosa: A systematic review and meta-analysis. *Current Opinion in Psychiatry, 32*, 445-455. doi: 10.1016/j.brat.2009.11.001

Berends, T., van Meijel, B., Nugteren, W., Deen, M., Danner, U.N., Hoek, H.W., & van Elberg, A. (2016). Rate, timing, and predictors of relapse in patients with anorexia nervosa following a relapse prevention program: a cohort study. *BMC Psychiatry, 16*, 316. DOI 10.1186/s12888-016-1019-y

Berg, K.C., Peterson, C.B., Frazier, P., & Crow, S. (2011). Convergence of scores on the interview and questionnaire versions of the Eating Disorder Examination: A meta-analytic review. *Psychological Assessment*, *23*, 714-724. doi: 10.1037/a0023246

Bjorntorp, P. (2002). Definition and classification of obesity. In C. G. Fairburn & K.D. Brownell (Eds.), Eating disorders and obesity (2nd ed.) (pp. 377-381). New York: Guilford.

Bouton, M.E. (2011). Learning and the persistence of appetite: Extinction and the motivation to eat and overeat. *Physiology & Behavior, 103*, 51-58. doi: 10.1016/j.physbeh.2010.11.025

Broomfield, C., Stedal, K., Touyz, S., & Rhodes, P. (2017). Labeling and defining severe and enduring anorexia nervosa: A systematic review and critical analysis. *International Journal of Eating Disorders, 50,* 611-623. https://doi.org/10.1002/eat.22715.

Brown, T.A., Holland, L.A., & Keel, P.K. (2014). Comparing operational definitions of DSM-5 anorexia nervosa for research contexts. *International Journal of Eating Disorders, 47*, 76-84. DOI: 10.1002/eat.22184.

Bruce, L. J., & Ricciardelli, L. A. (2016). A systematic review of the psychosocial correlates of intuitive eating among adult women. *Appetite, 96*, 454–472. doi:10.1016/j. Appet.2015.10.012

Bujang, M.A. & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Cohen's Kappa. *Epidemiology Biostatistics and Public Health*, *14*, e12267-1-e12267-10. DOI: 10.2427/12267.

Carter, P. I., & Moss, R. A. (1984). Screening for anorexia and bulimia nervosa in a college population. Problems and limitations. *Addictive Behaviors, 9*, 417–419. http://dx.doi.org/10.1016/0306-4603(84)90045-5.

Cash, T.F., Novy, P.L., & Grant, J.R. (1994). Why do women exercise? Factor analysis and further validation of the Reasons for Exercise Inventory. *Perceptual and Motor Skills, 78*, 539-544. DOI: 10.2466/pms.1994.78.2.539

Castel, L.D., Williams, K.A., Bosworth, H.B., Eisen, S.V., Hahn, E.A., Irwin, D.E., Kelly, M.A.R., Morse, J., Stover, A., DeWalt, D.A., & DeVellis, R.F. (2008). Content validity in the PROMIS social-health domain: A qualitative analysis of focus-group data. *Quality of Life Research, 17*, 737-749. DOI 10.1007/s11136-008-9352-3

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., DeVellis, R., DeWalt, D., Fries, J.F., Gershon, R., Hahn, E.A., Lai, J.-S., Pilkonis, P., Revicki, D., Rose, M., Weinfurt, K., & Hays, R. (2010). Initial adult health items banks and first wave testing of the Patient-Reported Outcomes Measurement Information System

(PROMIS) Network: 2005-2008. *Journal of Clinical Epidemiology, 63*, 1179-1194. doi: 10.1016/j.jclinepi.2010.04.011

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24,* 385-396. http://dx.doi.org/10.2307/2136404

Davis, H. A., Guller, L., & Smith, G. T. (2016). Developmental trajectories of compensatory exercise and fasting behavior across the middle school years. *Appetite, 107*, 330–338. https://doi.org/10.1016/j.appet.2016.08.098

Denny, K. N., Loth, K., Eisenberg, M. E., & Neumark-Sztainer, D. (2013). Intuitive eating in young adults. Who is doing it, and how is it related to disordered eating behaviors? *Appetite, 60*, 13–19. doi:10.1016/j.appet.2012.09.029

Denys, D. (2011). Obsessionality & compulsivity: A phenomenology of obsessive-compulsive disorder. *Philosophy, Ethics, and Humanities in Medicine, 6,* 1-7. doi: 10.1186/1747-5341-6-3

Eddy, K.T., Dorer, D.J., Franko, D.L., Tahilani, B.S., Thompson-Brenner, H., & Herzog, D.B. (2008). Diagnostic crossover in anorexia nervosa and bulimia nervosa: Implications for DSM-V. *American Journal of Psychiatry, 165*, 245-250. doi:10.1176/appi.ajp.2007.07060951

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160. DOI: 10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. DOI: 10.3758/BF03193146

Fairburn, C.G. & Beglin, S.J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370. DOI https://doi.org/10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-%23

Fairburn, C.G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509-528. https://doi.org/10.1016/S0005-7967(02)00088-8.

Fichter, M.M., Quadflieg, N., Hedlund, S. (2006). Twelve-year course and outcome predictors of anorexia nervosa. *International Journal of Eating Disorders, 39,* 87-100. doi: 10.1002/eat.20215

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). Structured clinical interview for DSM-IV axis I disorders, clinician version (SCID-CV). Washington, DC: American Psychiatric Press.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test. Psychometric features and clinical correlates. *Psychological Medicine, 12*, 871–878. http://dx.doi.org/10.1017/S0033291700049163.

Grilo, C.M., Masheb, R.M. & Wilson, G.T. (2001). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *Journal of Consulting and Clinical Psychology, 69*, 317-322. DOI: 10.1037/0022-006X.69.2.317

Halmi, K.A., Eckert, E., Marchi, P., Sampugnaro, V., Apple, R., & Cohen, J. (1990). Comorbidity of psychiatric diagnoses in anorexia nervosa. *Archives of General Psychiatry, 48*, 712- 718. doi:10.1001/archpsyc.1991.01810320036006

Hambrook, D., Oldershaw, A., Rimes, K., Schmidt, U., Tchanturia, K., Treasure, J., Richards, S., & Chalder, T. (2011). Emotional expression, self-silencing, and distress tolerance in anorexia nervosa and chronic fatigue syndrome. *British Journal of Clinical Psychology, 50*, 310-325. DOI:10.1348/014466510X519215

Harney, M.B., Fitzsimmons-Craft, E.E., Maldonado, C.R., & Bardone-Cone, A.M. (2014). Negative affective experiences in relation to stages of eating disorder recovery. *Eating Behaviors*, 15, 24-30. https://doi.org/10.1016/j.eatbeh.2013.10.016

HealthMeasures (2018, November 1*). Available PROMIS measures for adults*. Retrieved from http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-adult-measures

Hilbert, A., de Zwaan, M., & Braehler, E. (2012). How frequent are eating disturbances in the population? Norms of the Eating Disorder Examination-Questionnaire. PloS one, https://doi.org/10.1371/journal.pone.0029125

Hohlstein, L.A., Smith, G.T., & Atlas, J.G. (1998). An application of expectancy theory to eating disorders: Development and validation of measures of eating and dieting expectancies. *Psychological Assessment 10*, 49-58. DOI: 10.1037/1040-3590.10.1.49

Holtkamp, K.K., Müller, B.B., Heussen, N.N., Remschmidt, H.H., Herpertz-Dahlmann, B.B. (2005). Depression, anxiety, and obsessionality in long-term recovered patients with adolescent-onset anorexia nervosa. *European Child & Adolescent Psychiatry, 14*, 106–110. DOI: https://doi.org/10.1007/s00787-005-0431-5

Keel, P.K., Dorer, D.J., Franko, D.L., Jackson, S.C., & Herzog. D.B. (2005). Postremission predictors of relapse in women with eating disorders. *American Journal of Psychiatry, 162*, 2263-2268. https://doi.org/10.1176/appi.ajp.162.12.2263

Khalsa, S.S., Portnoff, L.C., McCurdy-McKinnon, D., & Feusner, J.D. (2017). What happens after treatment? A systematic review of relapse, remission, and recovery in anorexia nervosa. *Journal of Eating Disorders, 5,* 1-12. doi: 10.1186/s40337-017-0145-3

King, M. B. (1989). Eating disorders in general practice population. Prevalence, characteristics and follow-up at 12 to 18 months. *Psychological Medicine, Monograph Suppl., 14*, 1–34. http://dx.doi.org/10.1017/S0264180100000515.

King, M. B. (1991). The natural history of eating pathology in attenders to primary medical care. *International Journal of Eating Disorders, 10*, 379–387. http:// dx.doi.org/10.1002/1098-108X(199107)10:43.0.CO;2-I.

Lerman, D.C. & Iwata, B.A. (1995). Prevalence of the extinction burst and its attenuation during treatment. *Journal of Applied Behavior Analysis, 28*, 93-94. doi: 10.1901/jaba.1995.28-93

Radloff, L.S. (1991). The use of the Center for Epidemiological Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescence, 20*, 149-166. DOI: 10.1007/BF01537606

Radloff, L.S. & Locke, B.Z. (1986). The Community Mental Health Assessment Survey and the CES-D Scale. In Weissman, M., Myers, J., & Ross, C. (eds.), *Community Surveys*. Rutgers University Press, New Brunswick, NJ.

Radloff, L.S., & Teri., L. (1986). The use of the Center for Epidemiological Studies Depression Scale with older adults. In Brink, T.L. (ed.), *Clinical Gerontology: A Guide to Assessment and Intervention*. Haworth Press, New York.

Linville, D., Brown, T., Sturm, K., & McDougal, T. (2012). Eating disorders and social support: Perspectives of recovered individuals. *Eating Disorders, 20,* 216-231. https://doi.org/10.1080/10640266.2012.668480

Luce, K.H. & Crowther, J.H. (1999). The reliability of the Eating Disorder Examination-Self Report Questionnaire version (EDE-Q). *International Journal of Eating Disorders, 25*, 349-351. DOI https://doi.org/10.1002/(SICI)1098-108X(199904)25:3<349::AID-EAT15>3.0.CO;2-M

Luce, K.H., Crowther, J.H., & Pole, M. (2008). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders, 41,* 273-276. https://doi.org/10.1002/eat.20504

McKinley, N.M. & Hyde, J.S. (1996). The objectified body consciousness scale: Development and validation. *Psychology of Women Quarterly, 20*, 181-215. https://doi-org.libproxy.lib.unc.edu/10.1111/j.1471-6402.1996.tb00467.x

Meehan, K.G., Loeb, K.L., Roberto, C.A., & Attia, E. (2006). Mood change during weight restoration in patients with anorexia nervosa. *International Journal of Eating Disorders, 39*, 587-589, https://doi.org/10.1002/eat.20337

Metzger, R. L. (1976). A reliability and validity study of the State-Trait Anxiety Inventory. *Journal of Clinical Psychology, 32*, 276–278. http://dx.doi.org/10.1002/ 1097-4679(197604)32:23.0.CO;2-G.

Mischoulon, D., Eddy, K.T., Keshaviah, A., Dinescu, D., Ross, S.L., Kass, A.E., Franko, D.L., & Herzog, D.B. (2011). Depression and eating disorders: Treatment and course. *Journal of Affective Disorders, 130*, 470-477. doi:10.1016/j.jad.2010.10.043.

Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy, 44,* 53-62. https://doi.org/10.1016/j.brat.2004.12.003

Mond, J.M., Hay, P.J., Rodgers, B., Owen, C., & Beumont, P.J. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, *42*, 551-567. https://doi.org/10.1016/S0005-7967(03)00161-X

Moskowitz, L. & Weiselberg, E. (2017). Anorexia nervosa/atypical anorexia nervosa. Current Problems in Pediatric and Adolescent *Health Care, 47*, 70-84. https://doi.org/10.1016/j.cppeds.2017.02.003

Noetel, M., Dawson, L., Hay, P., & Touyz, S. (2017). The assessment and treatment of unhealthy exercise in adolescents with anorexia nervosa: A Delphi study to synthesize clinical knowledge. *International Journal of Eating Disorders, 50,* 378–388. https://doi.org/10.1002/eat.22657

Pearson, C. M., Combs, J. L., & Smith, G. T. (2010). A risk model for disordered eating in late elementary school boys. *Psychology of Addictive Behaviors, 24*, 696–704. https://doi.org/10.1037/a0020358

Pike, K.M. (1998). Long-term course of anorexia nervosa: Response, relapse, remission, and recovery. *Clinical Psychology Review, 18,* 447-475. https://doi.org/10.1016/S0272-7358(98)00014-2

Pollice, C., Kaye, W.H., Greeno, C.G., & Weltzin, T.E. (1997). Relationship of depression, anxiety, and obsessionality to the state of illness in anorexia nervosa. *International Journal of Eating Disorders, 21*, 367-376. https://doi.org/10.1002/(SICI)1098-108X(1997)21:4<367::AID-EAT10>3.0.CO;2-W

Pritchard, M.E. & Yalch (2009). Relationships among loneliness, interpersonal dependency, and disordered eating in young adults. *Personality and Individual Differences, 46*, 341-346. doi:10.1016/j.paid.2008.10.027

Quick, V. M., & Byrd-Bredbenner, C. (2013). Eating Disorders Examination Questionnaire (EDE-Q): Norms for US college students. *Eating and Weight Disorders, 18*, 29–35. https://doi.org/10.1007/s40519-013-0015-1

Radloff, L.S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401. https://doi.org/10.1177/014662167700100306

Ranstam, J. & Cook, J.A. (2017). Kaplan-Meier curve. *British Journal of Surgery, 104,* 442-442. https://doi.org/10.1002/bjs.10238

Richards, P.S., Crowton, S., Berrett, M.E., Smith, M.H., & Passmore, K. (2017). Can patients with eating disorders learn to eat intuitively? A 2-year pilot study. *Eating Disorders, 25*, 99-113. DOI: 10.1080/10640266.2017.1279907

Riley, E.N., Combs, J.L., Jordan C.E., & Smith, G.T. (2015). Negative urgency and lack of perseverance: Identification of differential pathways of onset and maintenance risk in the longitudinal prediction of non-suicidal self-injury. *Behavior Therapy, 46*, 439-448. doi:10.1016/j.beth.2015.03.002.

Rule, W. R., & Traver, M. D. (1983). Test-retest reliabilities of the State-Trait Anxiety Inventory in a stressful social analogue situation. *Journal of Personality Assessment, 47*, 276–277. http://dx.doi.org/10.1207/s15327752jpa4703_8.

Schaumberg, K., Jangmo, A., Thornton, L.M., Birgegard, A., Almqvist, C., Norring, C., Larsson, H., & Bulik, C.M. (2019). Patterns of diagnostic transition in eating disorders: A longitudinal population study in Sweden. *Psychological Medicine, 49,* 819-827. doi: 10.1017/S0033291718001472.

Silberstein, L.R., Stiegel-Moore, R.H., Timko, C., & Rodin, J. (1988). Behavioral and psychological implications of body dissatisfaction: Do men and women differ? *Sex Roles: A Journal of Research, 19*, 219-232. http://dx.doi.org/10.1007/BF00290156

Simmons, J.R., Smith, G.T., Hill, K.H. (2002). Validation of eating and dieting expectancy measures in two adolescent samples. *International Journal of Eating Disorders, 31*, 461-473. DOI https://doi org.libproxy.lib.unc.edu/10.1002/eat.10034

Simons, J.S. & Gaher, R.M. (2005). The Distress Tolerance Scale: Development and validation of a self-report measure. *Motivation and Emotion, 29*, 83-102. DOI: 10.1007/s11031-005-7955-3

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). STAI manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Steinhausen, H.-C. (2002). The outcome of anorexia nervosa in the 20th century. American Journal of Psychiatry, 159, 1284-1293. https://doi.org/10.1176/appi.ajp.159.8.1284

Strober, M., Freeman, R., & Morrell, W. (1997). The long-term course of severe anorexia nervosa in adolescents: Survival analysis of recovery, relapse, and outcome predictors over 10-15 years in a prospective study. *International Journal of Eating Disorders, 22*, 339-360. https://doi.org/10.1002/(SICI)1098-108X(199712)22:4<339::AID-EAT1>3.0.CO;2-N.

Taylor, S., Koch, W. J., & McNally, R. J. (1992). How does anxiety sensitivity vary across the anxiety disorders? *Journal of Anxiety Disorders, 6*, 249–259. http:// dx.doi.org/10.1016/0887-6185(92)90037-8.

Thomas, J.J., Roberto, C.A., & Brownell, K.D. (2009). Eighty-five per cent of what? Discrepancies in the weight cut-off for anorexia nervosa substantially affect the prevalence of underweight. *Psychological Medicine*, 39, 833-843. DOI: 10.1017/S0033291708004327

Tylka, T. L. (2006). Development and psychometric evaluation of a measure of intuitive eating. *Journal of Counseling Psychology, 53*, 226 –240. doi:10.1037/0022-0167.53.2.226

Tylka, T.L. & Kroon Van Diest, A.M. (2013). The Intuitive Eating Scale-2: Item Refinement and Psychometric Evaluation with College Women and Men. *Journal of counseling Psychology, 60*, 137-153. DOI: 10.1037/a0030893

Tylka, T. L., & Wilcox, J. A. (2006). Are intuitive eating and eating disorder symptomatology opposite poles of the same construct? *Journal of Counseling Psychology, 53*, 474–485. doi:10.1037/0022-0167.53.4.474

Viera, A.J. & Garrett, J.M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*, 360-363.

Villa, V., Manzoni, G.M., Pagnini, F., Castelnuovo, G., Cesa, G.L., & Molinari, E. (2009). Do coping strategies discriminate eating disordered individuals better than eating disorder features? An explorative study on female inpatients with anorexia and bulimia nervosa. *Journal of Clinical Psychology in Medical Settings, 16*, 297-303. DOI: 10.1007/s10880-009-9172-9

Walsh, B.T. (2013). The enigmatic persistence of anorexia nervosa. *American Journal of Psychiatry, 170,* 477-484. DOI: 10.1176/appi.ajp.2012.12081074

Watson, D., & Clark, L. A. (1984). Negative affectivity. The disposition to experience aversive emotional affects. *Psychological Bulletin, 55*, 465–490. http://dx.doi.org/ 10.1037/0033-2909.96.3.465.

Welch, E., Birgegard, A., Parling, T., & Ghaderi, A. (2011). Eating disorder examination questionnaire and clinical impairment assessment questionnaire: General population and clinical norms for young adult women in Sweden. *Behaviour Research and Therapy, 49*, 85-91. https://doi.org/10.1016/j.brat.2010.10.010

Whiteside, S.P. & Lynam, D.R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*, 669-689. https://doi.org/10.1016/S0191-8869(00)00064-7

Whiteside, S.P., Lynam, D.R., Miller, J.D., & Reynolds, S.K. (2005). Validation of the UPPS impulsive behavior scale: A four-factor model of impulsivity. *European Journal of Personality, 19*, 559-574. https://doi.org/10.1002/per.556

Wonderlich, S.A., Bulik, C.M., Schmidt, U., Steiger, H., & Hoek, H.W. (2020). Severe and enduring anorexia nervosa: Update and observations about the current clinical reality. *International Journal of Eating Disorders, 53*, 1303-1312. DOI: 10.1002/eat.23283.

Zimmerman, M., Posternak, M.A., & Chelminski, I. (2014). Using a self-report depression scale to identify remission in depressed outpatients. *American Journal of Psychiatry, 161*, 1911-1913. Doi: https://doi.org/10.1176/ajp.161.10.1911