



Published in final edited form as:

*Eat Behav.* 2021 August ; 42: 101540. doi:10.1016/j.eatbeh.2021.101540.

## An empirical evaluation of the diagnostic threshold between full-threshold and sub-threshold bulimia nervosa

Sarah N. Johnson,

Kelsie T. Forbush<sup>\*</sup>,

Trevor James Swanson,

Kara A. Christensen

University of Kansas, United States of America

### Abstract

Previous research has failed to find differences in eating disorder and general psychopathology and impairment between people with sub- and full-threshold bulimia nervosa (BN). The purpose of the current study was to test the validity of the distinction between sub- and full-threshold BN and to determine the frequency of objective binge episodes and inappropriate compensatory behaviors that would best distinguish between sub- and full-BN. Community-recruited adults (83.5% female) with current sub-threshold ( $n = 105$ ) or full-threshold BN ( $n = 99$ ) completed assessments of eating-disorder psychopathology, clinical impairment, internalizing problems, and drug and alcohol misuse. Receiver operating characteristic curve analysis was used to evaluate whether eating-disorder psychopathology, clinical impairment, internalizing problems, and drug and alcohol misuse could empirically discriminate between sub- and full-threshold BN. The frequency of binge episodes and inappropriate compensatory behaviors ( $AUC = 0.94$ ) was “highly accurate” in discriminating between sub- and full-threshold BN; however, only objective binge episodes was a significant predictor of BN status. Internalizing symptoms ( $AUC = 0.71$ ) were “moderately accurate” at distinguishing between sub- and full-BN. Neither clinical impairment ( $AUC = 0.60$ ) nor drug ( $AUC = 0.56$ ) or alcohol misuse ( $AUC = 0.52$ ) discriminated between groups. Results suggested that 11 episodes of binge eating and 17 episodes of inappropriate compensatory behaviors optimally distinguished between sub- and full-BN. Overall, results provided mixed support for the distinction between sub- and full-threshold BN. Future research to clarify the most meaningful way to discriminate between sub- and full-threshold is warranted to improve the criterion-related validity of the diagnostic system.

<sup>\*</sup>Corresponding author at: Department of Psychology, University of Kansas, 1415 Jayhawk Blvd., Lawrence, KS 66045, United States of America. kforbush@ku.edu (K.T. Forbush).

#### CRedit authorship contribution statement

The first (SJ) author conceptualized the research idea, wrote the draft of the manuscript, and conducted the statistical analyses. The corresponding author (KF) assisted with conceptualizing the research idea and writing the manuscript, and designed the study from which the data originated. The third author (TS) assisted with the statistical analyses and interpretation of the results, and the fourth author (KC) assisted with writing the manuscript. All authors contributed to and approved the final manuscript.

#### Declaration of competing interest

The authors declare that they have no conflict of interest.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eatbeh.2021.101540>.

## Keywords

Diagnosis; Classification; DSM-5; Bulimia nervosa; Other specified feeding and eating disorders (OSFED); ROC curve analysis

---

## 1. Introduction

According to the current *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, full-threshold bulimia nervosa (BN) and sub-threshold BN (diagnosed as other specified feeding and eating disorder) are distinct disorders (American Psychiatric Association, 2013). Although the diagnostic label of “sub-threshold” connotes a less severe eating disorder, the empirical basis for separate diagnostic classes for sub- and full-threshold eating disorders has not been supported by past meta-analytic research (Thomas et al., 2009). Previous studies have found that sub- and full-threshold eating disorders do not meaningfully differ on levels of clinical impairment, eating-disorder and general psychopathology, and genetic risk factors (Chapa et al., 2018; Fairweather-Schmidt & Wade, 2014; Wade & O’Shea, 2015), leading clinicians and researchers to question the utility of the distinction. If the distinction between sub- and full-threshold BN is not valid, then clinicians may have limited diagnostic information to inform treatment planning (e.g., length and intensity of treatment). Another consequence is that labeling EDs as sub-threshold may prevent people from receiving appropriate treatment due to lack of insurance coverage (Thompson & Park, 2016). Thus, it is necessary to improve the current diagnostic classification system to better patient outcomes.

Although past research has not found clinically meaningful differences between sub- and full-threshold BN when comparing the two groups (e.g., Chapa et al., 2018), another method of examining the extent to which differences between sub- and full-threshold BN exist is to test whether measures of psychopathology can empirically discriminate between sub- and full-threshold cases. Receiver operating characteristic (ROC) curve analysis is one method that can be utilized for this purpose. For example, past research has used ROC curve analysis to test the discriminatory accuracy of hypomania versus mania (Benazzi, 2009). ROC curve analysis quantifies how useful a measure is in terms of the amount information it provides across the full range of cut-off scores (McFall & Treat, 1999). ROC curve analysis can also determine if there is any threshold at which the disorders are different, rather than imposing a pre-selected cutoff value between groups. Determining a measure’s information value without a specific cutoff score is useful because a single cutoff score may not be appropriate across different clinical situations and, as a result, sensitivity and specificity can vary widely depending on the prevalence of a disorder in the population (McFall & Treat, 1999). Further, examining a measure’s performance across the range of scores yields information about the measure’s overall performance at all levels of the distribution, rather than accuracy at a single point.

Differences on measures of eating and general psychopathology are particularly relevant to identifying the appropriateness of the distinction between sub- and full-threshold BN. If the categorical distinction between sub- and full-threshold BN is valid, then these two groups

should differ on levels of eating pathology and related impairment. Further, given the high comorbidity between eating disorders and internalizing psychopathology (Hudson et al., 2007) and substance abuse (Hudson et al., 2007), studying differences in non-eating-disorder psychopathology between sub- and full-threshold BN is of high clinical interest.

The primary aim of the current study was to use ROC curve analysis to evaluate measures of eating-disorder impairment, internalizing problems, and drug and alcohol misuse with respect to their ability to discriminate between sub- and full-threshold BN cases. Based on past research (Chapa et al., 2018), we hypothesized that eating-disorder related impairment, internalizing symptoms, and drug and alcohol misuse would not provide an adequate level of informational value when distinguishing between groups. Of note, a portion of the data ( $n = 125$ ) from Chapa et al. (2018) are included in the current study; however, the data from Chapa et al. (2018) used different measures and analyses than the current study.

Our secondary aims were to evaluate how well ED-symptoms discriminated between groups and test which symptoms of eating-disorder and internalizing psychopathology were best for discriminating between groups. Given that sub- and full-threshold BN are diagnosed based on the frequency of binge episodes and inappropriate compensatory behaviors, we hypothesized that objective binge eating and inappropriate compensatory behaviors, such as purging, restricting, and excessive exercise, would provide statistically significant predictive power in distinguishing between sub-threshold and threshold BN cases. Lastly, a third, exploratory aim was to identify the ideal frequency criteria of objective binge episodes and inappropriate compensatory behaviors to optimally distinguish between sub- and full-BN.

## 2. Method

### 2.1. Participants and procedures

Participants were individuals from a large longitudinal study of eating-disorder psychopathology (see Forbush et al., 2018 for details) who met *DSM-5* criteria for full-threshold ( $n = 105$ ) or sub-threshold ( $n = 99$ ) BN. Individuals with full-threshold BN met all of the *DSM-5* criteria for BN; whereas, individuals with sub-threshold BN engaged in binge eating and inappropriate compensatory behaviors at a lower frequency. Participants were recruited from the general community, thus, the sample is a convenience sample; however, in the parent study, current and lifetime eating disorder diagnoses were similar to national representative data for BN (Bohrer et al., 2017; Hudson et al., 2007). Only baseline data were included in the current study.

Participant demographics by group are presented in Table 1. There were no statistically significant differences between sub- and full-threshold BN groups for age, BMI, sex, or proportion of ethnic and racial minority participants.

The Institutional Review Board approved all study procedures and informed consent was obtained prior to engaging in testing. Baseline procedures were completed in person. Following informed consent, participants completed semi-structured interviews and self-report measures. Height and weight were measured using a wall-mounted stadiometer and digital scale. Bachelor's and master's-level clinicians conducted the interviews and were

supervised by the corresponding author (KF) or an advanced PhD student during weekly diagnostic consensus meetings. Interviews were audiotaped with permission, and 10% of interviews were randomly selected to be rated by an independent interviewer to examine inter-rater reliability.

## 2.2. Measures

**2.2.1. The Eating Pathology Symptoms Inventory–Clinician Rated Version (EPSI-CRV)**—The EPSI-CRV (Forbush et al., 2020) is a semi-structured diagnostic interview that assesses dimensional eating disorder constructs and categorical ED diagnoses. The EPSI-CRV demonstrated strong convergent validity with SCID-I diagnoses and other eating disorder measures (Forbush et al., 2020) and excellent discriminant validity from depression and anxiety symptoms. In the current study, diagnoses of current BN showed evidence for excellent inter-rater reliability (Gwet’s AC1 of 0.85).

**2.2.2. Clinical Impairment Assessment Questionnaire (CIA)**—The CIA (Bohn & Fairburn, 2008; Bohn et al., 2008) is a 16-item self-report questionnaire that assesses impairment in intrapersonal, interpersonal, and cognitive functioning due to eating-disorder pathology. CIA scores range from 0 to 48, and the CIA has shown evidence for excellent test-retest reliability and convergent validity with clinicians’ ratings of impairment and other measures eating pathology (Bohn et al., 2008). In the current sample internal consistency was 0.90.

**2.2.3. Inventory of Depression and Anxiety Symptoms-II (IDAS-II)**—The IDAS-II (Watson et al., 2012) is a 99-item self-report questionnaire with 18 subscales that assess non-eating-disorder internalizing psychopathology. The IDAS-II has shown evidence for good-to-excellent two-week test-retest reliability, convergent validity with other measures of internalizing psychopathology, and criterion validity with *DSM-IV* diagnoses made by clinicians (Nelson et al., 2018; Watson et al., 2012). Internal consistency for each scale ranged from Cronbach’s alpha of 0.66 (Weight Gain) to 0.90 (Well-being, Claustrophobia, and Checking). Compared to national norms (Nelson et al., 2018), the current sample had negligible differences on Ill Temper, Claustrophobia, Checking, Mania, and Panic ( $d = 0.003$ – $0.11$ ), and small elevations were found on Traumatic Avoidance, Ordering, Insomnia, Suicidality, Euphoria, Traumatic Stress, and Cleaning ( $d = 0.20$ – $0.29$ ) subscales. Medium effects were found on Social Anxiety, Lassitude, Appetite Loss, and Well Being ( $d = 0.51$ – $0.61$ ), and large effects on Dysphoria and Appetite Gain ( $d = 0.75$ – $1.14$ ) subscales.

**2.2.4. Alcohol Use Disorders Identification Test (AUDIT)**—The AUDIT (Babor et al., 2001) was used to measure harmful or hazardous alcohol consumption. The AUDIT contains ten self-report items, and demonstrated evidence for acceptable construct validity and test-retest reliability in a variety of settings (de Meneses-Gaya et al., 2009). Audit scores range from 0 to 40 and a total score of 8 or more is consistent with harmful or hazardous drinking (Babor et al., 2001). In the current study, internal consistency was 0.86.

**2.2.5. Drug Abuse Screening Test (DAST)**—The DAST (Skinner, 1982), a 28-item self-report questionnaire, was used to measure abuse of drugs other than alcohol. Across

multiple studies, the DAST has shown evidence for moderate-to-high levels of test-retest reliability and validity for use in clinical populations (Yudko et al., 2007). DAST scores range from 0 to 10 and a total score of 6 provides excellent sensitivity as well as specificity in identifying individuals with substance use disorders (Skinner, 1982). Internal consistency in the current study was 0.78.

### 2.3. Data analyses

Separate logistic regressions were conducted for predicting sub- or full-threshold BN diagnosis using: 1) total frequency of objective binge episodes and inappropriate compensatory behaviors over the past three-months from the EPSI-CRV, 2) CIA, 3) all 18 scales of the IDAS-II, 4) AUDIT, and 5) DAST. Logistic regression models were evaluated using ROC curve analysis, a graphical plot of a measure's true positive rate (sensitivity) against its false positive rate (1 – specificity). Optimal values that maximize the average true and false positive rates are plotted, and the area under the curve (AUC) is calculated to examine the overall accuracy of the model. In other words, AUC evaluates each model's overall ability to discriminate between sub- and full-threshold cases. AUC represents the probability that a randomly selected participant with full-threshold BN would have a higher score on the measure than a randomly selected participant with sub-threshold BN. AUC values range from 0.50 to 1.00, with values of 0.50 indicating that the model is performing no better than random chance, 0.51–0.70 as “less accurate,” 0.71–0.90 as “moderately accurate,” 0.91–0.99 as “highly accurate,” and 1.00 as perfect (Swets, 1988). Analyses were performed in R (Version 3.6.1).

Due to missing values for all scale items, the final sample for the EPSI-CRV model was 193 and the CIA model was 198. Missing values analysis revealed that less than 1% of individual items on the CIA were missing. Little's test of Missing Completely at Random (MCAR) was not significant,  $\chi^2(105, N = 198) = 87.04, df = 145, p = .898$ , indicating that data were MCAR and that multiple imputation would be appropriate. The “Amelia” package in R was used to impute missing values for the CIA using 1000 bootstrapped resamples (Honaker et al., 2011).

## 3. Results

Means and standard deviations by sub-group are presented in Table 1. The model containing the CIA was not a significant predictor of BN status,  $R_{McFadden}^2 = 0.05$ , although the CIA model fit better than the intercept-only model,  $\chi^2(1, N = 198) = 6.71, p = .010$ . The CIA fell in the “less accurate” range of predicting BN status with AUC = 0.61, Optimal Sensitivity = 0.61, and Optimal Specificity = 0.64,  $R_{McFadden}^2 = 0.15$  (see Fig. 1b). The model containing the IDAS-II fell in the “moderately accurate” range in predicting BN status, AUC = 0.70, Optimal Sensitivity = 0.62, Optimal Specificity = 0.72,  $R_{McFadden}^2 = 0.14$  (see Fig. 2a). The IDAS-II model did not fit better than the intercept-only model,  $\chi^2(18, N = 198) = 25.47, p = .113$ . Only the Appetite Gain subscale was a significant predictor of BN status,  $\beta = 0.21$  (0.07),  $p = .003$ .

Neither the AUDIT nor the DAST were accurate in predicting BN status,  $AUC = 0.52$ , Optimal Sensitivity = 0.32, Optimal Specificity = 0.77,  $R_{McFadden}^2 = 0.06$  and  $AUC = 0.56$ , Optimal Sensitivity = 0.46, Optimal Specificity = 0.70,  $R_{McFadden}^2 = 0.10$ , respectively (see Fig. 2b and c). Neither the AUDIT model,  $\chi^2(1, N = 193) = 0.47, p = .494$ , nor the DAST model,  $\chi^2(1, N = 185) = 2.07, p = .150$ , fit better than the intercept-only model.

The model containing frequencies of binge eating and inappropriate compensatory behaviors was a good predictor of BN status,  $R_{McFadden}^2 = 0.34$  according to McFadden's criteria (Louviere et al., 2000). The full model was a significantly better fit to the data than the intercept-only model,  $\chi^2(3, N = 193) = 81.51, p < .001$ . The model containing binge episode and inappropriate compensatory behavior frequencies of the EPSI-CRV fell in the "highly accurate" range in predicting BN status with  $AUC = 0.94$ , Optimal Sensitivity = 0.95, and Optimal Specificity = 0.89 (see Fig. 1a). Only objective binge episodes were a significant predictor of BN status,  $\beta = 0.10 (0.02), p < .001$  (see Table 2). This log-odds value corresponds to an expected increase of 52.5% in the odds being classified as threshold BN (relative to sub-threshold BN) given a one-episode increase in objective binge episodes. Of note, a model with the self-report Eating Pathology Symptoms Inventory (EPSI; Forbush et al., 2013) measure found that the EPSI fell in the "less accurate" range ( $AUC = 0.68$ ) and only the Binge Eating subscale was a significant predictor of BN status,  $\beta = 0.09 (0.03), p = .002$ , corroborating the results of the EPSI-CRV model.

To determine the optimal frequency of objective binge episodes and inappropriate compensatory behaviors to discriminate between sub- and full-threshold BN, we examined three separate models evaluating 1) objective binge episodes, 2) inappropriate compensatory behaviors (i.e., purging, excessive exercise, restricting as defined as a concrete period of time without eating or eating considerably less than others of the same, age, sex, and weight), and 3) purging and excessive exercise (i.e., inappropriate compensatory behaviors excluding restricting). A model with only purging and excessive exercise was examined to understand how the optimal frequency of inappropriate compensatory behaviors would change if restricting were not included given that some researchers may use narrower definitions of fasting (e.g., 8 or more waking hours without eating anything) compared to the definition employed in the current study. According to the objective binge episodes model, 11 objective binge episodes in the past three-months was the optimal frequency to distinguish between sub- and full-threshold BN. Given that the total episodes of inappropriate compensatory behaviors were highly skewed (skewness = 3.32) outliers above two standard deviations were excluded from the model. After removing outliers, the model with all inappropriate compensatory behaviors identified 17 episodes of inappropriate compensatory behaviors in the past three-months as the ideal frequency to discriminate between sub- and full-threshold BN, and the model with only excessive exercise and purging identified 13 episodes.

#### 4. Discussion

The primary aim of the current study was to use ROC curve analysis to evaluate criterion validity of the sub- and full-threshold BN distinction. We evaluated measures of eating-disorder impairment, internalizing problems, and drug and alcohol misuse with respect to



their ability to discriminate between sub- and full-threshold BN cases. We hypothesized that neither clinical impairment (CIA) internalizing symptoms (IDAS-II), alcohol use (AUDIT), and substance use (DAST) would not discriminate between groups. Overall, results supported our hypotheses. The AUC values for the CIA fell in the “less accurate” range, indicating that the eating disorder-related clinical impairment did not adequately discriminate between sub- and full-threshold BN cases. This suggests that although level of impairment discriminates non-ED pathology from clinically significant levels of eating pathology (Bohn et al., 2008), it is less useful at discriminating among eating-disorder severity levels. Results, together with past studies (e.g., Chapa et al., 2018; Fairweather-Schmidt & Wade, 2014; Wade & O’Shea, 2015), support the claim that the distinction between sub- and full-threshold BN may not be meaningful in terms of the impairment individuals experience due to their eating disorder. Our hypothesis that internalizing psychopathology and alcohol and substance use would not discriminate between BN cases was supported. The AUC values for the IDAS-II fell on the lower end of the “moderately accurate” range, and The AUDIT and DAST’s AUC values were close to 0.50, indicating that they only perform slightly better than random chance. Taken together, our results suggested that sub- and full-threshold BN groups were more similar than different on levels of non-eating disorder psychopathology.

A secondary aim was to test the current DSM criteria for symptom frequency. Our hypothesis that the frequency of binge episodes and inappropriate compensatory behaviors would provide an acceptable level of predictive power in discriminating between groups was supported. The model using frequency of objective binge episodes and inappropriate compensatory behaviors fell in the “highly accurate” range, suggesting that the frequency of binge episodes and inappropriate compensatory behaviors accurately discriminated between sub- and full-BN. This finding makes conceptual sense, given that by definition, individuals with full-threshold BN engage in more episodes of objective binge eating and inappropriate compensatory behaviors than their sub-threshold counterparts.

An additional secondary aim of this study was to test which symptoms of eating-disorder and internalizing psychopathology best discriminated between groups. We hypothesized that objective binge eating and inappropriate compensatory behavior would be significant predictors of group status. In line with our hypotheses, objective binge episodes was a significant predictor of BN status and Cohen’s  $d$  between sub- and full-threshold BN was 0.91. However, contrary to expectation, use of inappropriate compensatory behaviors (i.e., purging, restricting, and excessive exercise frequency) was non-significant in the model. Taken together, our results suggest that the variability in BN case status was driven by objective binge episodes, indicating that the current binge eating frequency criterion is closer to the optimum criterion than the inappropriate compensatory behavior frequency criterion. To improve the utility of the diagnostic system it may be more important to focus on revising the frequency criterion for inappropriate compensatory behaviors. In terms of internalizing symptoms and alcohol and substance use, only the Appetite Gain subscale of the IDAS-II was a significant predictor of BN group and the only statistically significant different IDAS-II subscale ( $d = 0.45$ ). There was a significant correlation between objective binge episodes and the Appetite Gain subscale,  $r(188) = 0.32$ ,  $p < .001$ , and the significance of the Appetite Gain subscale is most likely explained by this overlap in content with binge eating.

Results suggest internalizing and externalizing symptoms do not discriminate between sub- and full-BN.

The third aim was to identify the ideal frequency criteria of objective binge episodes and inappropriate compensatory behaviors to optimally distinguish between sub- and full-BN. The results of this study suggested that lowering the frequency of objective binge episodes from 12 to 11 for full-threshold BN would be optimal for distinguishing between sub- and full-threshold BN. Although the frequency criterion for binge eating and compensatory behaviors were lowered from 24 episodes in *DSM-IV* to 12 episodes in *DSM-5*, the results from this study, and others (Chapa et al., 2018; Fairweather-Schmidt & Wade, 2014; Wade & O'Shea, 2015), suggest that the frequency criterion for binge episodes may need to be further lowered to more validly capture clinically significant BN cases. The ideal cutoff for inappropriate compensatory behaviors was 17, suggesting that the frequency criterion for inappropriate compensatory behaviors may need to be increased. Given the accumulating evidence that the diagnostic threshold between sub- and full-threshold BN has questionable validity, future research in this area is warranted. If the findings of the current study are replicated using different methods, the DSM should consider identifying a more appropriate frequency criterion for binge episodes and inappropriate compensatory behaviors that could meaningfully differentiate between levels of BN pathology.

The results of the current study indicate that the frequency criteria for objective binge episodes and inappropriate compensatory behaviors may differ from each other. This could suggest that individuals may be engaging in multiple compensatory behaviors following a binge episode or could be compensating in response to a subjective binge episode. Although remembering different cut-offs may increase clinician burden and require a more sensitive assessment, blanket modifications to both criteria may not meaningfully capture differences between sub and full-BN. In other areas of psychopathology, such as PTSD and bipolar disorder, clinicians have evidenced their ability to remember different symptom frequency cut-offs; thus, we expect that clinicians could successfully apply different cut-offs for BN.

It is important to note the limitations of the current study. The majority of participants were White and female, which limits the generalizability of findings to other racial/ethnic groups and gender identities. Future research should examine ROC curve analysis in diverse samples, particularly given emerging evidence that clinical impairment may differ according to gender (Richson et al., 2021). Second, the current study only examined whether sub- and full-threshold BN groups differed on eating pathology, eating-disorder-related impairment, and non-eating-disorder psychopathology. Sub- and full-threshold BN groups could differ on other important metrics, such as medical risk, long-term course, and response to treatment. Given the absence of many variables related to course and outcome that are relevant to treatment decisions, future research should continue to evaluate the ideal frequency of eating disorder symptoms prior to changing diagnostic criteria. Lastly, given that new data were not collected for this study, a priori power was not computed as a means of determining sample size. Future research would benefit from calculating power ahead of data collection to better assess the reliability of our findings.



Future research could also explore better ways to categorize and represent eating-disorder severity. For example, dimensional models of eating pathology do not impose arbitrary cut-offs between sub-threshold and full-threshold disorders (Forbush et al., 2018; Wildes & Marcus, 2013). Dimensional models may better capture information on eating-disorder psychopathology and severity than the current sub-threshold and full-threshold distinction, particularly given that groups did not differ in clinical impairment or general psychopathology and likely have a similar need for treatment. The *International Classification of Disease*, 11th Revision (ICD-11) has proposed to broaden the diagnosis of full-threshold BN to include subjective binge eating (Uher & Rutter, 2012). Some individuals in our sample who were diagnosed with sub-threshold BN according to DSM-5 criteria would be classified as full-threshold BN according to the proposed ICD-11 criteria. It will be important for future research to examine if the proposed change to the BN diagnosis better distinguishes severity levels of those with sub- and full-threshold BN. Moreover, BN severity ratings are only determined by the frequency of inappropriate compensatory behaviors. Given previous research has failed to find meaningful support for the current severity ratings (Gianini et al., 2017; Grilo et al., 2015) and the current study's findings suggesting that the variability in BN case status was largely driven by binge episodes, future research should evaluate alternative severity rating systems. For example, the use of binge eating or a combination binge eating and inappropriate compensatory behaviors may provide more valid and informative severity ratings than the current system. Future research should also explore how different definitions of fasting and restricting affect the optimal cutoff for inappropriate compensatory behaviors. The current study employed a definition of restricting consistent with the Eating Disorder Examination (Fairburn et al., 2008). Restriction was defined as a concrete period of time without eating or eating considerably less than others of the same, age, sex, and weight. Importantly, this definition encompasses a larger set of behaviors used to compensate for food intake than definitions of fasting that only assess a portion of the day (e.g., 8 h a day). However, more strict definitions of fasting may find a lower optimal cutoff. The current study also did not examine the threshold of eating disordered behaviors that distinguishes between sub-threshold BN and no eating disorder. Future research should examine the optimal frequency of binge eating and compensatory behaviors which distinguish between a clinically significant eating disorder and disordered eating.

In conclusion, objective binge-eating episodes significantly discriminated between sub- and full-threshold BN cases, whereas inappropriate compensatory behaviors, clinical impairment, drug, and alcohol misuse did not distinguish between groups. Internalizing symptoms significantly discriminated between groups, but this was mainly due to the presence of appetite gain, which had conceptual overlap with binge eating. Despite differences in objective binge-eating episodes, we found that individuals with sub- or full-threshold BN were more alike than dissimilar on indicators other than binge eating. Finally, the results of our study suggested that a reduced number of binge-eating episodes from 12 to 11 and an increased number of inappropriate compensatory behaviors from 12 to 17 best discriminated between groups.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

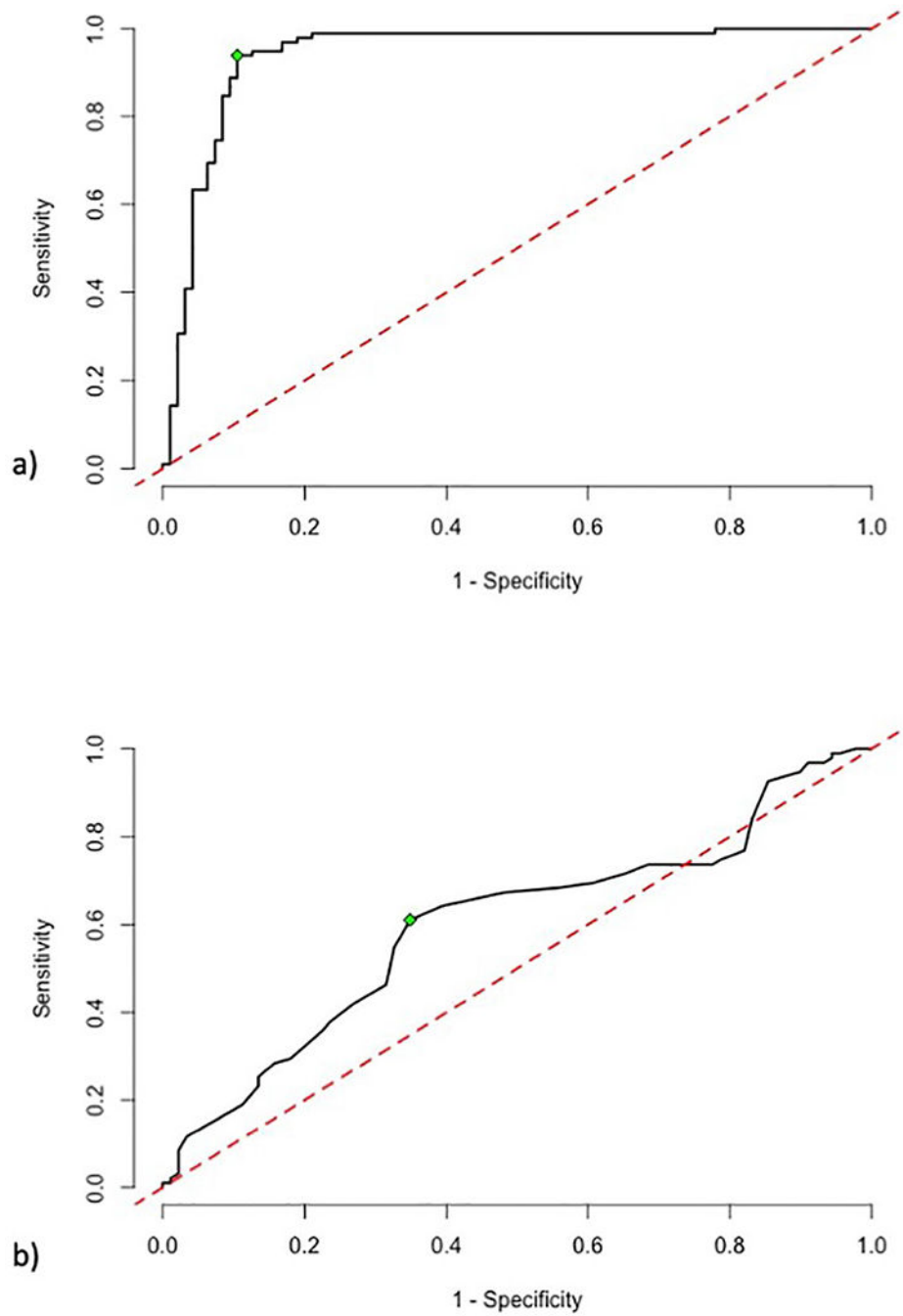
## Financial support

This work was supported by a Clifford B. Kinley Trust Award, University of Kansas Research Excellence Initiative Grant, and University of Kansas New Faculty Research Grant awarded to Kelsie T. Forbush, PhD. KAC is funded by a TL1 postdoctoral fellowship awarded by Frontiers: University of Kansas Clinical and Translational Science Institute (#TL1TR002368) through a CTSA grant from NCATS. Any opinions, conclusions, recommendations, or other statements expressed in this work are those of the authors and do not necessarily reflect the views of the funding agencies.

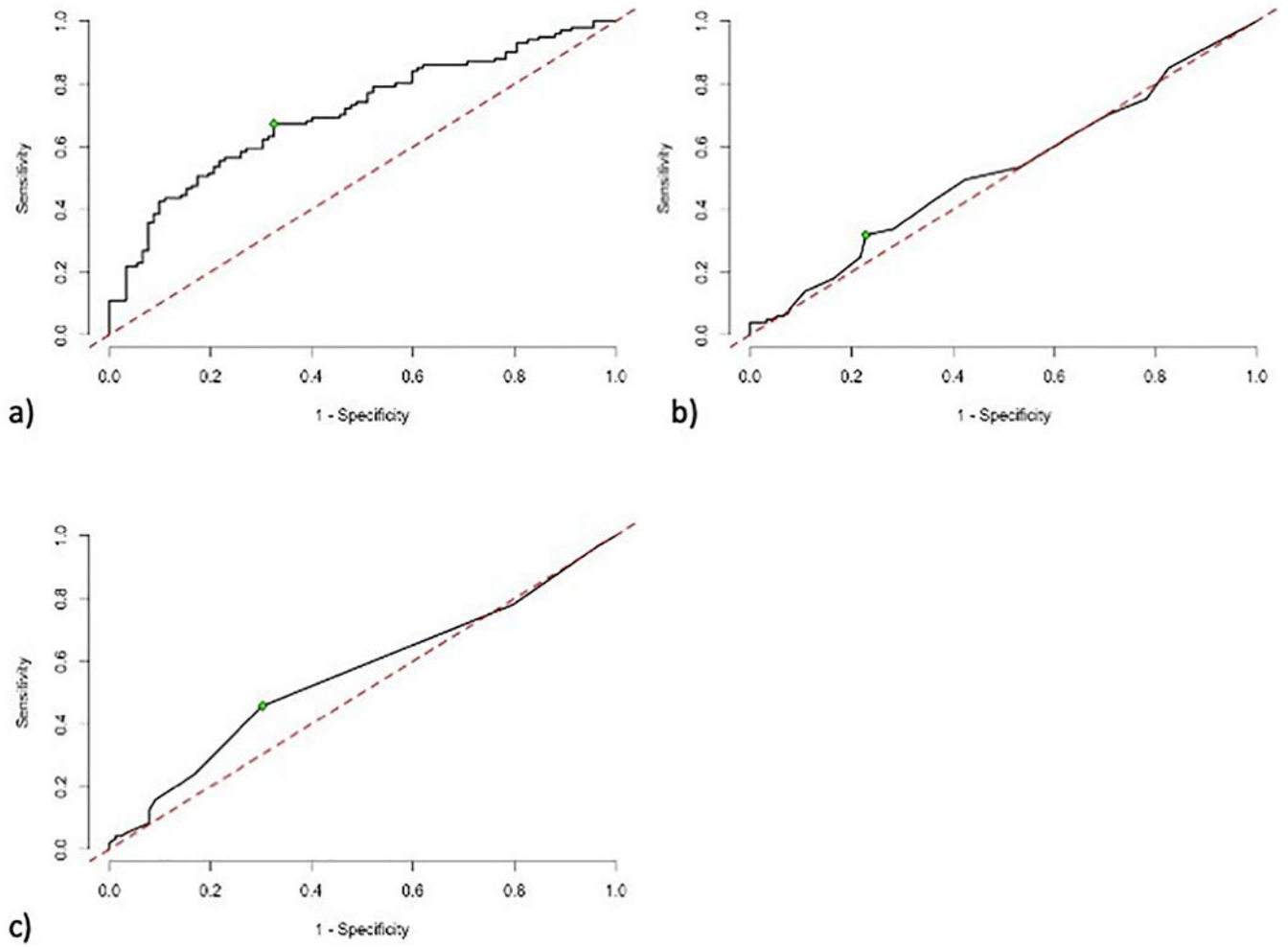
## References

- Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- Babor T, Higgins-Biddle JC, Saunders JB, & Monteiro MG (2001). *The Alcohol Use Disorders Identification Test: Guidelines for use in primary care* (pp. 1–40). Geneva: World Health Organization.
- Benazzi F (2009). What is hypomania? Tetrachoric factor analysis and kernel estimation of DSM-IV hypomanic symptoms. *Journal of Clinical Psychiatry*, 70(11), 1514–1521. 10.4088/JCP.09m05090. [PubMed: 19744407]
- Bohn K, Doll HA, Cooper Z, O'Connor M, Palmer RL, & Fairburn CG (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour Research and Therapy*, 46(10), 1105–1110. 10.1016/j.brat.2008.06.012. [PubMed: 18710699]
- Bohn K, & Fairburn CG (2008). *Clinical Impairment Assessment Questionnaire (CIA 3.0)*. Cognitive behavioral therapy for eating disorders. New York: Guilford Press.
- Bohrer BK, Carroll IA, Forbush KT, & Chen P-Y (2017). Treatment seeking for eating disorders: Results from a nationally representative study. *International Journal of Eating Disorders*, 50(12), 1341–1349. 10.1002/eat.22785. [PubMed: 28963793]
- Chapa DAN, Bohrer BK, & Forbush KT (2018). Is the diagnostic threshold for bulimia nervosa clinically meaningful? *Eating Behaviors*, 28, 16–19. 10.1016/j.eatbeh.2017.12.002. [PubMed: 29245042]
- de Meneses-Gaya C, Zuairi AW, Loureiro SR, & Crippa JAS (2009). Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. *Psychology & Neuroscience*, 2(1), 83–97. 10.3922/j.psns.2009.1.12.
- Fairburn CG, Cooper ZE, & O'Connor M (2008). *Eating disorder examination* (edition 16.0D). In *Cognitive behavior therapy and eating disorders* (pp. 265–308). Guilford Press.
- Fairweather-Schmidt AK, & Wade TD (2014). DSM-5 eating disorders and other specified eating and feeding disorders: Is there a meaningful differentiation? *International Journal of Eating Disorders*, 47(5), 524–533. 10.1002/eat.22257. [PubMed: 24616045]
- Forbush KT, Bohrer BK, Hagan KE, Chapa DANN, Perko V, Richson BN, ... Wildes JE (2020). Development and initial validation of the eating pathology symptoms inventory – Clinician rated version (EPSI - CRV). *Psychological Assessment*, 32(10), 943–955. 10.1037/pas0000820. [PubMed: 32718162]
- Forbush KT, Chen P-Y, Hagan KE, Chapa DAN, Gould SR, Eaton NR, & Krueger RF (2018). A new approach to eating-disorder classification: Using empirical methods to delineate diagnostic dimensions and inform care. *International Journal of Eating Disorders*. 10.1002/eat.22891.
- Forbush KT, Wildes JE, Pollack LO, Dunbar D, Luo J, Patterson K, ... Watson, D. D. of P. S. (2013). Development and validation of the eating pathology symptoms inventory (EPSI). *Psychological Assessment*, 25(3), 859–878. 10.1037/a0032639. [PubMed: 23815116]
- Gianini L, Roberto CA, Attia E, Walsh BT, Thomas JJ, Eddy KT, ... Sysko R (2017). Mild, moderate, meaningful? Examining the psychological and functioning correlates of DSM-5 eating disorder

- severity specifiers. *International Journal of Eating Disorders*, 50(8), 906–916. 10.1002/eat.22728. [PubMed: 28489323]
- Grilo CM, Ivezaj V, & White MA (2015). Evaluation of the DSM-5 severity indicator for bulimia nervosa. *Behaviour Research and Therapy*, 67, 41–44. 10.1016/j.brat.2015.02.002. [PubMed: 25744910]
- Honaker J, King G, & Blackwell M (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45, 1–47.
- Hudson JI, Hiripi E, Pope HG, & Kessler RC (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry*, 61(3), 348–358. 10.1016/j.biopsych.2006.03.040. [PubMed: 16815322]
- Louviere JJ, Hensher DA, & Swait JD (2000). “Frontmatter”, in *stated choice methods: Analysis and applications*. Cambridge University Press.
- McFall RM, & Treat TA (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*, 50(1), 215–241. 10.1146/annurev.psych.50.1.215.
- Nelson GH, O’Hara MW, & Watson D (2018). National norms for the expanded version of the inventory of depression and anxiety symptoms (IDAS-II). *Journal of Clinical Psychology*, 74(6), 953–968. 10.1002/jclp.22560. [PubMed: 29327338]
- Richson BN, Johnson SN, Swanson TJ, Chistensen KA, Forbush KT, & Wildes JE (2021). Predicting probable eating disorder case-status in men using the Clinical Impairment Assessment: Evidence for a gender-specific threshold. *Eating Behaviors*. In press.
- Skinner HA (1982). The drug abuse screening test. *Addictive Behaviors*, 7(4), 363–371. 10.1016/0306-4603(82)90005-3. [PubMed: 7183189]
- Swets JA (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857), 1285–1293. 10.1126/science.3287615. [PubMed: 3287615]
- Thomas JJ, Vartanian LR, & Brownell KD (2009). The relationship between eating disorder not otherwise specified (EDNOS) and officially recognized eating disorders: Meta-analysis and implications for DSM. *Psychological Bulletin*, 135(3), 407–433. 10.1037/a0015326. [PubMed: 19379023]
- Thompson C, & Park S (2016). Barriers to access and utilization of eating disorder treatment among women. *Archives of Women’s Mental Health*, 19(5), 753–760. 10.1007/s00737-016-0618-4.
- Uher R, & Rutter M (2012). Classification of feeding and eating disorders: Review of evidence and proposals for ICD-11. In, Vol. 11. *World psychiatry* (pp. 80–92). Blackwell Publishing Ltd. 10.1016/j.wpsyc.2012.05.005. Issue 2.
- Wade TD, & O’Shea A (2015). DSM-5 unspecified feeding and eating disorders in adolescents: What do they look like and are they clinically significant? *International Journal of Eating Disorders*, 48(4), 367–374. 10.1002/eat.22303. [PubMed: 24854848]
- Watson D, O’Hara MW, Naragon-Gainey K, Koffel E, Chmielewski M, Kotov R, ... Ruggero CJ (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*, 19(4), 399–420. 10.1177/1073191112449857. [PubMed: 22822173]
- Wildes JE, & Marcus MD (2013). Incorporating dimensions into the classification of eating disorders: Three models and their implications for research and clinical practice. *The International Journal of Eating Disorders*, 46(5), 396–403. 10.1002/eat.22091. [PubMed: 23658078]
- Yudko E, Lozhkina O, & Fouts A (2007). A comprehensive review of the psychometric properties of the drug abuse screening test. *Journal of Substance Abuse Treatment*, 32 (2), 189–198. 10.1016/j.jsat.2006.08.002. [PubMed: 17306727]



**Fig. 1.** Receiver operating characteristic curve analysis showing the area under the curve for prediction of bulimia nervosa status using the behavioral frequencies of the Eating Pathology Symptoms Inventory-Clinician Rated Version (a) and Clinical Impairment Assessment (b).



**Fig. 2.** Receiver operating characteristic curve analysis showing the area under the curve for prediction of bulimia nervosa status using the Inventory of Depression and Anxiety Symptoms (a), Alcohol Use Disorders Screening Test (b), Drug Abuse Screening Test (c).

**Table 1**

Means, percentages, and standard deviations for sub- and full-threshold BN Groups.

	Sub-threshold BN <i>n</i> = 105	Full-threshold BN <i>n</i> = 99	<i>F</i>	<i>p</i>	Cohen's <i>d</i>
Age	23.86 (7.55)	24.40 (9.40)	0.20	0.655	0.06
BMI	26.43 (7.27)	26.83 (6.94)	0.16	0.687	0.06
% Female	77.55% (76)	89.32% (92)			
% White	73.40% (69)	75.73% (78)			
% Asian	15.96% (15)	17.48% (18)			
% African American	6.38% (6)	5.83% (6)			
% Native American/Alaskan Native	2.13% (2)	2.91% (3)			
% Multi-racial	5.32% (5)	5.83% (6)			
% Other race	6.4% (6)	0.98% (1)			
% Hispanic	10.42 (10)	7.77% (8)			
EPSI-CRV (past 3 months)					
Objective Binge Episodes	9.04 (22.73)	35.86 (34.78)	41.01	<0.001***	0.91
Subjective Binge Episodes	5.90 (13.10)	1.35 (13.73)	5.83	0.017*	-0.34
Restricting Days	31.63 (26.22)	35.89 (25.66)	1.36	0.246	0.17
Purging Episodes	14.43 (30.26)	36.63 (81.46)	6.31	0.013*	0.36
Excessive Exercise Episodes	10.08 (16.80)	16.50 (25.18)	4.46	0.036*	0.30
Clinical Impairment Assessment	24.43 (8.84)	27.53 (9.23)	5.40	0.021*	0.34
IDAS					
Dysphoria	25.37 (8.12)	26.76 (7.82)	1.47	0.226	0.18
Lassitude	15.82 (5.22)	16.14 (5.23)	0.18	0.668	0.06
Insomnia	13.53 (5.43)	13.97 (5.55)	0.31	0.581	0.08
Suicidality	7.47 (2.90)	7.54 (2.65)	0.04	0.847	0.03
Appetite Loss	6.87 (2.75)	6.67 (3.13)	0.21	0.645	-0.07
Appetite Gain	8.72 (2.36)	9.88 (2.76)	9.83	0.010**	0.45
Well-being	18.03 (6.27)	18.51 (5.75)	0.31	0.578	0.08
Ill Temper	7.92 (3.3)	8.38 (3.88)	0.75	0.388	0.13
Mania	9.54 (4.80)	9.62 (4.85)	0.01	0.908	0.02
Euphoria	7.95 (3.33)	7.97 (3.37)	0.003	0.959	0.01
Panic	11.91 (3.82)	13.05 (5.66)	2.62	0.107	0.23
Social Anxiety	13.43 (5.32)	14.06 (6.27)	0.55	0.459	0.11
Claustrophobia	7.29 (3.75)	7.12 (3.94)	0.10	0.753	-0.05
Traumatic Intrusions	7.64 (3.29)	7.22 (3.61)	0.72	0.397	-0.12
Traumatic Avoidance	8.61 (3.73)	8.62 (3.90)	0.001	0.978	0.0
Checking	6.24 (3.11)	6.74 (3.51)	1.10	0.295	0.15
Ordering	9.65 (4.19)	9.68 (4.90)	0.002	0.963	0.01
Cleaning	9.90 (4.67)	10.77 (5.65)	1.35	0.248	0.17
AUDIT	5.85 (5.37)	6.42 (6.16)	0.46	0.497	0.10
% Above Clinical Cutoff	21.21% (21)	27.82% (32)			



	<b>Sub-threshold BN <i>n</i> = 105</b>	<b>Full-threshold BN <i>n</i> = 99</b>	<i>F</i>	<i>p</i>	<b>Cohen's <i>d</i></b>
DAST	3.55 (1.98)	4.04 (2.65)	2.02	0.158	0.21
% Above Clinical Cutoff	7.07% (7)	11.43% (12)			

\*  
 $p < .05$ .

\*\*  
 $p < .01$ .

\*\*\*  
 $p < .001$ .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Results of the Regressions with Sub- or Full-Threshold BN as the Outcome Variable.

	$\beta$	(S.E.)	<i>p</i>	AUC
Model 1				0.94
Objective Binge Episodes	0.10	0.02	<0.001 ***	
Inappropriate Compensatory Behaviors	0.007	0.004	0.07	
Model 2				0.61
Clinical Impairment Assessment		0.02	0.01 **	
Model 3				0.71
Dysphoria	0.03	0.04	0.50	
Lassitude	-0.02	0.04	0.56	
Insomnia	0.03	0.03	0.38	
Suicidality	0.03	0.07	0.63	
Appetite Loss	0.02	0.07	0.79	
Appetite Gain	0.21	0.07	0.004 **	
Well-being	0.05	0.04	0.23	
Ill Temp	-0.06	0.06	0.37	
Mania	0.06	0.05	0.19	
Euphoria	0.04	0.08	0.59	
Panic	0.08	0.05	0.10	
Social Anxiety	0.02	0.04	0.66	
Claustrophobia	-0.06	0.05	0.27	
Traumatic Intrusions	-0.10	0.06	0.11	
Traumatic Avoidance	0.009	0.06	0.88	
Checking	0.12	0.08	0.12	
Ordering	-0.09	0.06	0.11	
Cleaning	0.06	0.04	0.14	
Model 4				0.52
Alcohol use Disorders Identification Test	0.02	0.03	0.50	
Model 5				0.56
Drug Abuse Screening Test	0.09	0.07	0.16	

\*\*  
*p* < .01.\*\*\*  
*p* < .001.