Psychological Medicine

cambridge.org/psm

The natural course of binge-eating disorder: findings from a prospective, community-based study of adults

Kristin N. Javaras[1,2], Victoria F. Franco[1], Boyu Ren[1,2], Cynthia M. Bulik[3,4,5], Scott J. Crow[6,7], Susan L. McElroy[8,9], Harrison G. Pope Jr[1,2] and James I. Hudson[1,2]

[1]McLean Hospital, Belmont, MA, USA; [2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [3]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [5]Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [6]Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA; [7]Accanto Health, Saint Paul, MN, USA; [8]Lindner Center of HOPE, Mason, OH, USA and [9]Department of Psychiatry & Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH, USA

Original Article

Cite this article: Javaras KN, Franco VF, Ren B, Bulik CM, Crow SJ, McElroy SL, Pope Jr HG, Hudson JI (2024). The natural course of binge-eating disorder: findings from a prospective, community-based study of adults. *Psychological Medicine* **54**, 2906–2916. https://doi.org/10.1017/S0033291724000977

Received: 19 May 2023
Revised: 2 March 2024
Accepted: 3 April 2024
First published online: 28 May 2024

Keywords:
binge-eating disorder; eating disorders; epidemiology; machine learning; natural course; outcomes; predictors; relapse; remission

Corresponding author:
Kristin N. Javaras;
Email: kjavaras@mclean.harvard.edu

Abstract

Background. Epidemiological data offer conflicting views of the natural course of binge-eating disorder (BED), with large retrospective studies suggesting a protracted course and small prospective studies suggesting a briefer duration. We thus examined changes in BED diagnostic status in a prospective, community-based study that was larger and more representative with respect to sex, age of onset, and body mass index (BMI) than prior multi-year prospective studies.

Methods. Probands and relatives with current DSM-IV BED ($n = 156$) from a family study of BED ('baseline') were selected for follow-up at 2.5 and 5 years. Probands were required to have BMI > 25 (women) or >27 (men). Diagnostic interviews and questionnaires were administered at all timepoints.

Results. Of participants with follow-up data ($n = 137$), 78.1% were female, and 11.7% and 88.3% reported identifying as Black and White, respectively. At baseline, their mean age was 47.2 years, and mean BMI was 36.1. At 2.5 (and 5) years, 61.3% (45.7%), 23.4% (32.6%), and 15.3% (21.7%) of assessed participants exhibited full, sub-threshold, and no BED, respectively. No participants displayed anorexia or bulimia nervosa at follow-up timepoints. Median time to remission (i.e. no BED) exceeded 60 months, and median time to relapse (i.e. sub-threshold or full BED) after remission was 30 months. Two classes of machine learning methods did not consistently outperform random guessing at predicting time to remission from baseline demographic and clinical variables.

Conclusions. Among community-based adults with higher BMI, BED improves with time, but full remission often takes many years, and relapse is common.

Introduction

Existing epidemiological data support conflicting views (Keski-Rahkonen, 2021; Wilfley, Wilson, & Agras, 2003) of the natural course of binge-eating disorder (BED), the most prevalent eating disorder in the United States (Hudson, Hiripi, Pope, & Kessler, 2007; Taylor, Caldwell, Baser, Faison, & Jackson, 2007; Udo & Grilo, 2018). Community-based studies using retrospective assessments, including three nationally representative samples, suggest a protracted course for BED, with mean durations ranging from 7 to 16 years (Hudson et al., 2007; Pope et al., 2006; Taylor et al., 2007; Udo & Grilo, 2018). By contrast, several small prospective, community-based studies potentially suggest a more transient course (Brown, Klein, & Keel, 2015). For example, in an eight-year longitudinal study (Stice, Marti, & Rohde, 2013) of incident eating disorder cases in girls ages 12–15 years at baseline, 14 of 15 (93%) participants who met DSM-5 criteria for BED at some point during the study had gone at least a month without meeting full criteria for BED by 2-year follow-up. (Results for DSM-IV BED in the same sample were similar [Stice, Marti, Shaw, & Jaconis, 2009]). In a longitudinal study (Fairburn, Cooper, Doll, Norman, & O'Connor, 2000) of 48 women with DSM-IV BED who were age 16–35 years at baseline, seven (18%) of the 40 participants with follow-up data met criteria for an eating disorder at 5-year follow-up, including four (10%) participants who still had BED. In a study (Cachelin et al., 1999) of 31 women with DSM-IV BED and a mean age of 31.8 years, 11 (52%) of the 21 participants with follow-up data met full criteria for DSM-IV BED at six months follow-up, and the remaining 10 (48%) experienced less than

one binge-eating episode/week. However, no participants experienced more than two months without binge-eating episodes at either 3-month or 6-month follow-up.

All of these prior studies have limitations: the retrospective studies are subject to recall bias, whereas the prospective studies have been extremely small, excluded males, and either lacked longer-term follow-up (Cachelin et al., 1999) or included participants who were not representative of individuals with BED in terms of age of onset or BMI (Fairburn et al., 2000; Stice et al., 2013). More specifically, mean age of onset was approximately 17 years in Fairburn et al. (2000) and Stice et al. (2013), which is earlier than the 23–25 years reported in large nationally representative, community-based studies (Kessler et al., 2013; Taylor et al., 2007; Udo & Grilo, 2018). Similarly, of 36 participants with BED and recorded BMI in Fairburn et al. (2000), only 22% had a BMI of 30 or greater, in contrast to 68% of participants with 12-month BED in the most recent nationally representative study (Udo & Grilo, 2018). Thus, larger multi-year prospective studies that include men, middle-aged adults, and more individuals with higher BMI are needed to better understand the natural course of BED.

Further, little is known about the demographic and clinical predictors of the natural course of BED. One study (Cachelin et al., 1999), which used multiple linear regression to identify predictors of binge-eating frequency at 3- and 6-months follow-up in 21 participants, found that older age predicted greater frequency of binge eating at 3-month, but not 6-month, follow-up.

Thus, in this study, our primary aim was to characterize the course of BED, including time to remission and time to relapse after remission, based on prospective data from a longitudinal study (Hudson et al., 2010) of adults with BED. Our secondary aim was to apply machine learning methods to examine whether common, easily collected demographic and clinical measures from baseline could be used to predict time to remission. Note that the goal of prediction is fundamentally different from the goal of 'explanation,' or identification of causal structure (Diemer, Hudson, & Javaras, 2021; Yarkoni & Westfall, 2017).

Methods

Participants

At baseline, 156 individuals (99 probands and 57 relatives) with current DSM-IV BED were selected for a longitudinal follow-up study (Hudson et al., 2010) from among participants in a family study of BED (Hudson et al., 2006). Previous investigations using the family study data have examined the (retrospectively assessed) age of onset and diagnostic stability of BED (Pope et al., 2006), the familial aggregation and heritability of BED (Hudson et al., 2006; Javaras et al., 2008a), and the co-occurrence of psychiatric and other medical disorders with BED (Javaras et al., 2008b); in addition, a previous investigation using the longitudinal study data has examined the onset of metabolic syndrome components in BED (Hudson et al., 2010). Notably, probands in the family study were required to have body mass index (BMI) greater than 25 kg/m^2 (women) or 27 kg/m^2 (men), because a major aim of that study was to examine the familial aggregation of BED controlling for familial aggregation of higher BMI.

Of the 156 individuals selected for longitudinal follow-up, 19 (12 probands and 7 relatives) could not be assessed at 2.5-year follow-up, and an additional eight participants (5 probands and 3 relatives) could not be assessed at 5-year follow-up, corresponding to follow-up rates of 87.8% (2.5 years) and 82.7% (5 years), respectively.

Procedures and materials

Participants completed interviews and questionnaires at baseline and at 2.5-year and 5-year follow-up timepoints. Baseline assessments were conducted from October 2002 to July 2004. The longitudinal study protocol specified that the 2.5- and 5-year follow-up assessments would occur 24–36 months and 54–66 months, respectively, after the baseline assessment. There were three deviations from this timeline (two participants interviewed at 36.1 and 36.3 months for the 2.5-year follow-up, and one participant interviewed at 71.5 months for the 5-year follow-up).

Participants were interviewed by the same of three psychiatrists (including JIH and HGP) at all three timepoints. At each timepoint, the interviewer administered the Structured Clinical Interview for DSM-IV (SCID-IV) (First, Spitzer, Gibbons, & Williams, 2001), which included assessment of BED diagnostic status at that timepoint. At follow-up timepoints, the interviewer also gathered detailed information on any changes in BED symptoms since the last assessment, using questions based on the SCID-IV, together with the participants' retrospective chronology of these changes, which was recorded to the nearest month using a timeline that was reviewed with the participant. These assessments yielded information on: (a) BED diagnostic status (i.e. full DSM-IV BED, four sub-threshold forms of BED, and no BED; see Table 1) at each timepoint; and (b) changes in BED diagnostic status throughout the follow-up interval, along with the time at which the changes occurred, provided that the new status lasted for at least 3 months. The baseline SCID-IV information on other lifetime psychiatric disorders was used to create baseline predictors of time to remission from BED.

Additionally, at baseline, participants completed several questionnaires tapping constructs relevant to BED, including the Eating Disorder Inventory–2 (EDI-2; Garner, 1991), the Three Factor Eating Questionnaire (Stunkard & Messick, 1985), and the Barratt Impulsiveness Scale–11 (Patton, Stanford, & Barratt, 1995). Subscale scores for these questionnaires were used as predictors of time to remission from BED.

Table 1. Definitions of BED diagnostic status[a]

BED[b]	Meets full DSM-IV criteria for BED (including eating binges that average ⩾ 2 days a week)
Sub-threshold BED A	Meets all DSM-IV criteria for BED except for criterion B and/or C (associated features)
Sub-threshold BED B[b]	Eating binges that average ⩾ 1 day a week, but < 2 days a week
Sub-threshold BED C	Eating binges that average ⩾ 1 day a month, but < 1 day a week
Sub-threshold BED D	Eating 'mini-binges' that (a) do not fully meet DSM-IV criterion A for amount of food consumed or loss of control, but where the amount of food was at least *somewhat* large, and there was at least *some* loss of control, and (b) average ⩾ 2 days a week
No BED	No BED and no sub-threshold BED

Abbreviation: BED, binge-eating disorder.
[a]A (change in) status was recorded only if the new status was experienced for 3 or more months.
[b]Meets criteria for DSM-5 BED.

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

This study was approved by the Mass General Brigham Institutional Review Board, and all participants provided informed consent after the study procedures had been fully explained.

Analyses

We used R version 4.3.2 to perform all data analyses.

Below, remission refers to the first occurrence of no BED within the follow-up period, and relapse refers to the first occurrence of sub-threshold or full BED within the follow-up period following remission.

To graphically display how BED diagnostic status differed at the three assessment timepoints (baseline, 2.5 years, and 5 years), we used the ggplot function from the *ggalluvial* package. To visualize time to remission and time to relapse after remission, we used the survfit function from the *survival* package to calculate Kaplan–Meier curves with confidence intervals that account for clustering by family. To investigate the relationship between individual variables (e.g. duration of BED) and time to remission, we used Cox proportional hazards models (i.e. Cox regression) with robust variance estimation that accounted for clustering of observations by family.

For the main machine learning analyses, we used 38 demographic and clinical variables (see online Supplemental Materials Table S1) from the baseline timepoint to predict the time to remission for the 137 participants with follow-up data. The predictors were selected based on statistical and, more importantly, practical considerations: our goal was to pose a well-defined prediction problem using common and easily collected demographic and clinical measures from baseline as predictors of time to remission. We considered four machine learning methods (standard Cox regression; penalized Cox regression with lasso or elastic-net penalties; and random survival forests), which represent a combination of regression-based and tree-based methods similar to those we have used previously (Hudson et al., 2023), and that have been used in a prior study predicting treatment outcomes in a randomized controlled trial in BED (Forrest, Ivezaj, & Grilo, 2023). Because 1.0% of the predictor values were missing, we first created five imputed datasets. Then, for each imputed dataset, we performed 100 replicates of 5-fold cross-validation with pre-selected seeds to examine the performance of the four methods. We tuned hyper-parameters for penalized Cox regression and random survival forests via nested 5-fold cross-validation. To obtain a summary of our two chosen performance statistics (the concordance index and R^2 based on Schmid's robust estimate of prediction error [Rahman, Ambler, Choodari-Oskooei, & Omar, 2017; Schmid, Hielscher, Augustin, & Gefeller, 2011]), we then calculated the mean of the 500 performance statistics (5 folds times 100 replicates) for each method for a given imputed dataset, and then took the median of the mean performance statistics for each method across imputations (Marshall, Altman, Holder, & Royston, 2009; Wood, Royston, & White, 2015). We also performed six sensitivity analyses. More details of the machine learning analyses are provided in the online Supplemental Materials.

Results

Sample characteristics

Table 2 presents summary statistics for participants with follow-up data ($n = 137$). For these participants, the mean reported age was 47.2 years (standard deviation [S.D.] = 13.2); 78.1% reported their sex as female, and 21.9% reported their sex as male; and 87.6%, 11.7%, and 0.7% reported identifying as White non-Hispanic, Black non-Hispanic, and White Hispanic, respectively. Gender identity was not assessed. Regarding education, 23.4% reported completing high school (or equivalent) or less education, 31.4% reported completing some college, 17.5% reported completing a 2- or 4-year college, 10.9% reported completing some graduate or professional school, and 16.8% reported completing graduate or professional school. Information on income or net assets was not available. The mean duration of BED was 22.3 years (S.D. = 15.4), and the mean BMI was 36.1 kg/m^2 (S.D. = 7.5). Also, only a small proportion of participants ($n = 8$, 5.8%) had a history of DSM-IV anorexia nervosa (excluding the amenorrhea criterion) or bulimia nervosa at baseline.

Natural course of BED

We first present statistics calculated from participants' diagnostic status at the baseline and 2.5- and 5-year follow-up timepoints (see Fig. 1). The percentage of participants with full BED decreased over time, and the percentages with sub-threshold BED and no BED increased over time. Specifically, at the 2.5-year follow-up, 61.3%, 23.4%, and 15.3% of assessed participants had full BED, sub-threshold BED, and no BED, respectively, and, at the 5-year follow-up, 45.7%, 32.6%, and 21.7% of assessed participants had full BED, sub-threshold BED, and no BED, respectively. Notably, there was no diagnostic cross-over (e.g. from BED to anorexia nervosa or bulimia nervosa) at either of the follow-up timepoints. Additionally, for participants assessed at both follow-up timepoints, 35.0% of those experiencing remission (i.e. no BED) at 2.5-year follow-up had relapsed to either full or sub-threshold BED at 5-year follow-up.

We next present statistics calculated from participants' diagnostic status transition data across the 5-year follow-up period (see Fig. 2). The median time to remission (i.e. no BED) exceeded 60 months. The median time to relapse (i.e. to full BED or sub-threshold BED) after remission was approximately 30 months.

Predictors of time to remission

In the main analyses and the sensitivity analyses, none of the four machine learning methods had a summary measure of concordance better than random guessing for both performance statistics (see online Supplemental Materials for method performance in the main and sensitivity analyses).

Similarly, in standard Cox regression models for time to remission as a function of each baseline predictor (see online Supplemental Materials Table S1), no predictors would survive correction for multiple comparisons.

Discussion

We examined the natural course of BED, including time to remission and relapse, over a 5-year follow-up period for a community-based sample of men and women with BED (by DSM-IV criteria) whose ages ranged from 19 to 74 years and whose BMI ranged from 22.8 to 63.4 kg/m^2. We also used two classes of machine learning methods to predict time to remission as a function of baseline demographic and clinical variables.

In our sample, BED status improved from baseline to 2.5-year follow-up, and from 2.5-year follow-up to 5-year follow-up.

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

Table 2. Summary statistics for baseline and 2.5- and 5-year follow-up variables

Variable	Summary Statistics[a]		
Baseline variables:			
Sex			
Female	107	(78.1%)	
Male	30	(21.9%)	
Race/ethnicity			
White non-Hispanic	120	(87.6%)	
White Hispanic	1	(0.7%)	
Black non-Hispanic	16	(11.7%)	
Age (years)	47.2	(13.2)	[19–74]
Education			
Seventh to twelfth Grade, without high school degree	7	(5.1%)	
Completed high school or equivalent	25	(18.2%)	
Partial college	43	(31.4%)	
Completed 2-year college	11	(8.0%)	
Completed 4-year college	13	(9.5%)	
Partial graduate or professional school	15	(10.9%)	
Completed graduate or professional school	23	(16.8%)	
Marital status			
Married	63	(46.0%)	
Widowed	6	(4.4%)	
Divorced	29	(21.2%)	
Separated	5	(3.6%)	
Never married	34	(24.8%)	
Current BMI (kg/m^2)	36.1	(7.5)	[22.8–63.0]
BMI suppression[b] (kg/m^2)	3.2	(4.1)	[−0.6–22.0]
Lowest BMI since 18 years	23.7	(4.6)	[14.6–41.0]
BED duration (years)	22.3	(15.4)	[1–65]
BED family history[c] (proportion with BED)	0.34	(0.32)	[0.00–1.00]
Past anorexia nervosa[d]			
Absent	135	(98.5%)	
Present	2	(1.5%)	
Past bulimia nervosa[e]			
Absent	130	(94.9%)	
Present	7	(5.1%)	
Any lifetime substance use disorder[f]			
Absent	91	(66.4%)	
Present	46	(33.6%)	
Any lifetime major anxiety disorder[g]			
Absent	80	(58.4%)	
Present	57	(41.6%)	
Lifetime post-traumatic stress disorder			
Absent	123	(89.8%)	
Present	14	(10.2%)	

(*Continued*)

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

Table 2. (*Continued.*)

Variable	Summary Statistics[a]		
Lifetime obsessive-compulsive disorder			
Absent	127	(92.7%)	
Present	10	(7.3%)	
Any lifetime major mood disorder[h]			
Absent	57	(41.6%)	
Present	80	(58.4%)	
Barratt Impulsiveness Scale – 11			
Attentional Impulsiveness	14.5	(4.0)	[0–23]
Motor Impulsiveness	21.7	(4.8)	[0–32]
Nonplanning Impulsiveness	19.6	(5.9)	[2–34]
Eating Disorder Inventory – 2			
Drive for Thinness	8.8	(5.4)	[0–20]
Bulimia	7.3	(4.3)	[0–17]
Body Dissatisfaction	19.5	(7.0)	[0–27]
Ineffectiveness	6.1	(6.6)	[0–30]
Perfectionism	5.8	(4.5)	[0–18]
Interpersonal Distrust	3.3	(3.6)	[0–15]
Interoceptive Awareness	5.4	(4.9)	[0–21]
Maturity Fears	4.2	(3.7)	[0–19]
Three Factor Eating Questionnaire			
Dietary Restraint	7.6	(4.6)	[0–22]
Disinhibition	13.3	(2.0)	[8–16]
Hunger	10.2	(3.2)	[1–14]
2.5-Year follow-up variables:			
Assessment timing (months since baseline)	31.2	(2.7)	[25.9–36.3]
5-Year follow-up variables:			
Assessment timing (months since baseline)	61.4	(2.1)	[54.4–71.5]

Abbreviations: BED, binge-eating disorder; BMI, body mass index.
[a]Summary statistics presented for $n = 137$ participants who contributed to follow-up analyses. For continuous variables, summary statistics are mean, (standard deviation), [range]. For categorical variables, summary statistics are number and (percent) belonging to each category.
[b]BMI suppression defined as highest lifetime BMI – current BMI.
[c]BED family history defined as the proportion of the participant's interviewed relatives who met criteria for (full) DSM-IV BED based on the SCID-IV at baseline.
[d]Past anorexia nervosa defined as meeting criteria for DSM-IV anorexia nervosa (excluding the amenorrhea criterion).
[e]Past bulimia nervosa defined as meeting criteria for DSM-IV bulimia nervosa.
[f]Any lifetime substance use disorder defined as meeting criteria for any of the following lifetime DSM-IV disorders based on the SCID-IV: alcohol abuse; alcohol dependence; drug abuse; drug dependence.
[g]Any lifetime major anxiety disorder defined as meeting criteria for any of the following lifetime DSM-IV disorders based on the SCID-IV: panic disorder without agoraphobia; panic disorder with agoraphobia; agoraphobia without panic disorder; social phobia; generalized anxiety disorder.
[h]Any lifetime major mood disorder defined as meeting criteria for any of the following lifetime DSM-IV disorders based on the SCID-IV: bipolar I disorder; bipolar II disorder; major depressive disorder.

However, almost two-thirds (61.3%) of participants assessed at 2.5-year follow-up and almost half (45.7%) of participants assessed at 5-year follow-up still met criteria for full (DSM-IV) BED at those follow-up timepoints. Moreover, the percentages meeting the broader DSM-5 BED criteria at these follow-up timepoints would be even higher (i.e. 67.9% and 58.9%, respectively). Further, only a minority of individuals assessed at 2.5-year follow-up (15.3%) and 5-year follow-up (21.7%) experienced remission at those timepoints, respectively. Similarly, the diagnostic status transition data suggested that the median time to remission exceeds 5 years. Taken together, these results suggest that, under naturalistic circumstances, the majority of individuals with BED will experience some improvement in binge eating within a 5-year period, but achieving full remission within this time frame is not common – at least for prevalent cases of BED in adults with generally higher BMI. These results are consistent with findings from community-based studies using retrospective lifetime interviews (Hudson et al., 2007; Pope et al., 2006; Taylor, 2015; Udo & Grilo, 2018), including the family study from which our participants derived (Pope et al., 2006), all of which suggest that BED tends to have a multi-year course. These results are less consistent with findings from two prior

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

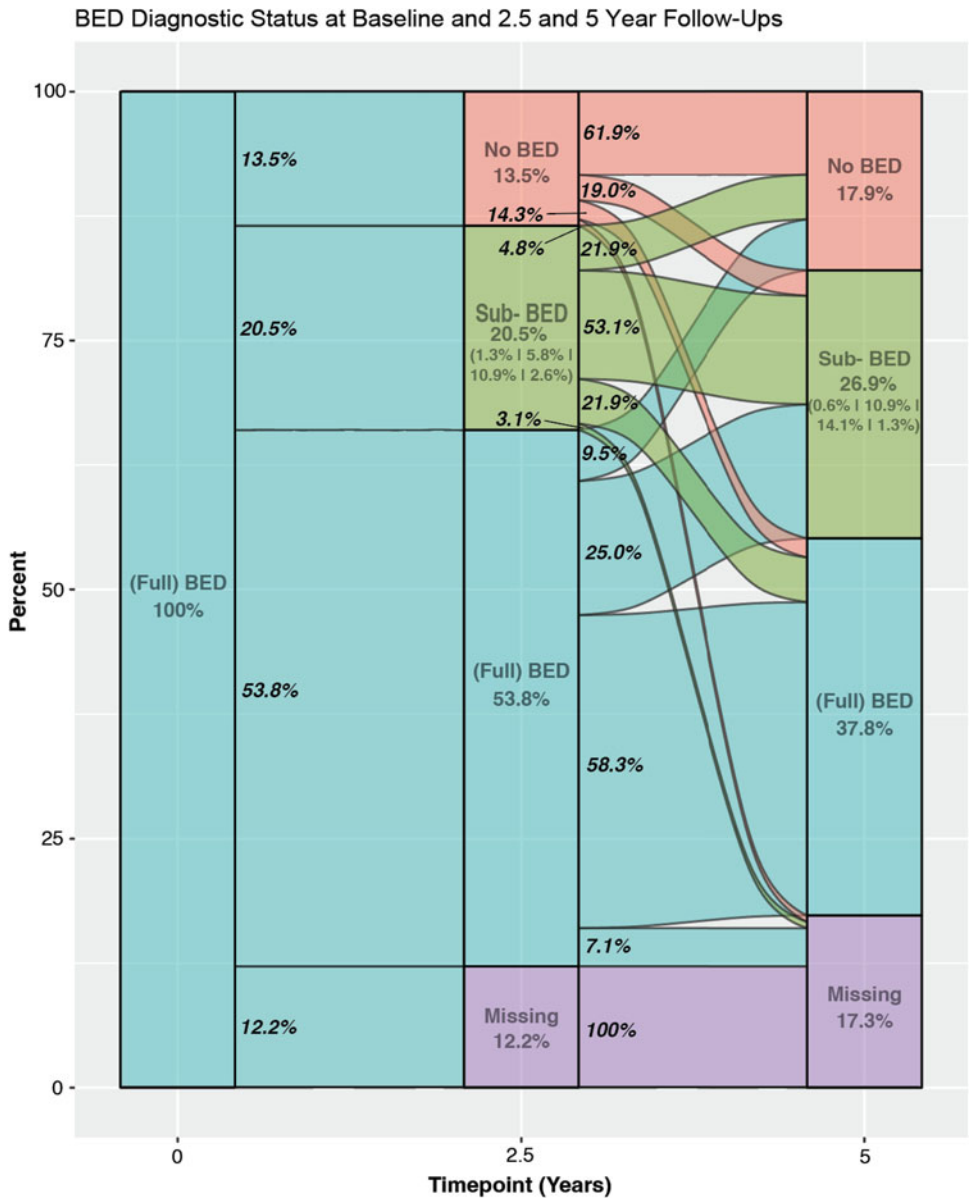

Figure 1. Alluvial plot for diagnostic status at baseline and 2.5- and 5-year follow-up timepoints. The figure illustrates diagnostic status at each timepoint, as well as transitions among these statuses. The non-italicized numbers refer to timepoint percentages (i.e. out of all 156 participants, what percentage belonged to each diagnostic status at a given timepoint); for the sub-threshold BED status, the numbers in the parentheses further represent the percentage with each type of sub-threshold BED (i.e. [A | B | C | D]). The italicized numbers refer to transition percentages (i.e. out of all participants with a given diagnostic status at a given time-point, what percentage transitioned to each diagnostic status at the next timepoint). Abbreviations: BED, binge-eating disorder; sub-BED, sub-threshold binge-eating disorder.

prospective, community-based studies with multi-year follow-up periods (Fairburn et al., 2000; Stice et al., 2013, 2009), as well as those from two multi-year prospective, community-based studies that followed adolescent girls (Glazer et al., 2019) or pregnant young adult women (Knoph et al., 2013) with BED-like presentations based on questionnaires that assessed some, but not all, BED criteria. Although different definitions of remission preclude fully comparing these prospective studies' results to our own, these prior studies suggest that BED is more transient, with few individuals meeting full criteria for BED at follow-up. In contrast to our study, which sampled prevalent cases from among adults with higher BMI or at familial risk for higher BMI (that is, having a proband relative with higher BMI), the prior prospective studies of BED focused on incident cases (Stice et al., 2013, 2009) and/or on adolescent (Stice et al., 2013, 2009) or young adult (Fairburn et al., 2000) females only, which may account for the more favorable prognosis in these other samples – especially given some evidence that younger age predicts better natural outcomes (Cachelin et al., 1999). Also of note, outcomes in our naturalistic study were generally worse than 4-year outcomes in a longitudinal follow-up study that recruited participants from both specialty clinics and the community (Agras, Crow, Mitchell, Halmi, & Bryson, 2009). Of the 104 participants with DSM-IV BED at baseline in this latter study, 82% no longer met criteria for an eating disorder at some point during the 4-year follow-up period, although a high proportion of these experienced a subsequent relapse (i.e. meeting criteria for any eating disorder). Notably, 70% of participants with BED reported

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

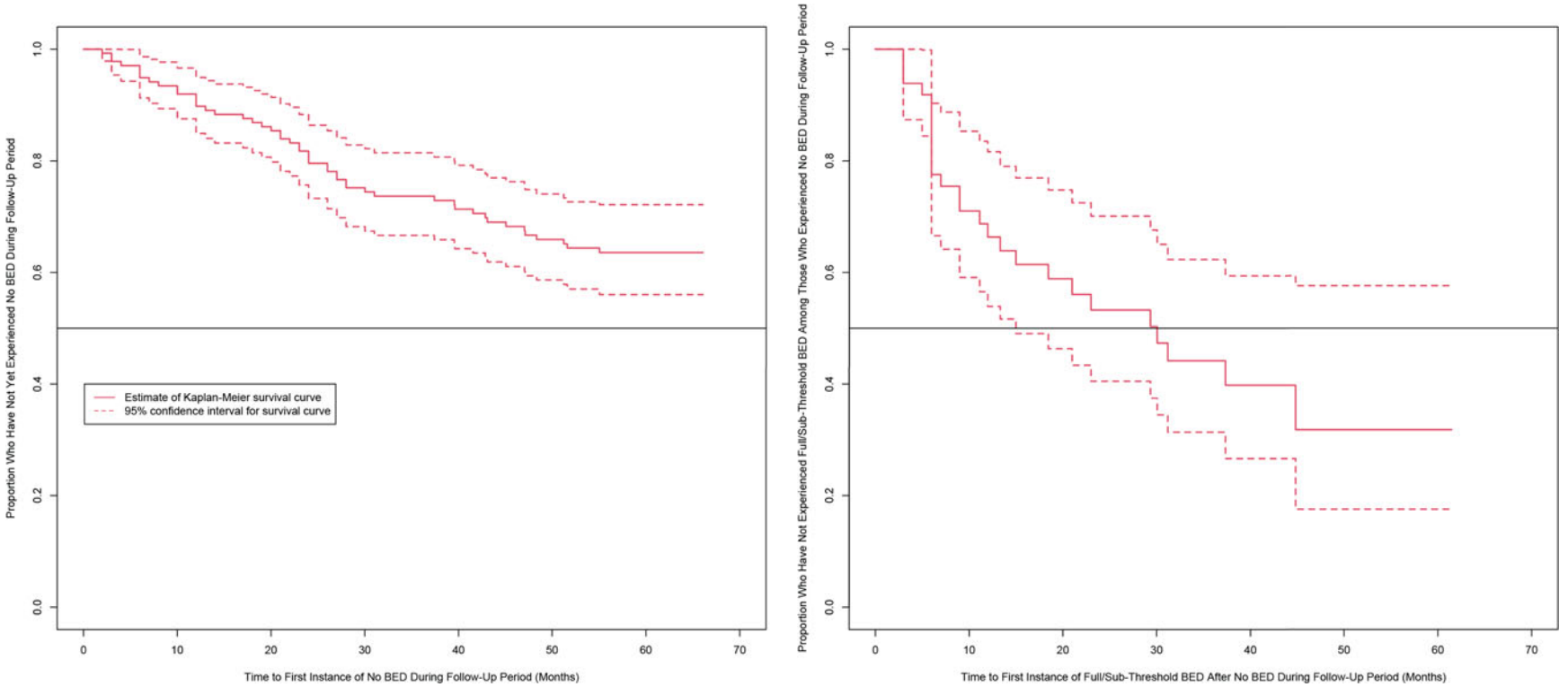

Figure 2. Survival curves for time to remission and time to relapse after remission during the 5-year follow-up period. Remission is defined as not having either (full) BED or sub-threshold BED for at least 3 months. Relapse is defined as having either (full) BED or sub-threshold BED for at least 3 months after experiencing remission. In the plots, the solid gray line is the Kaplan–Meier survival curve, and the dotted gray lines are the corresponding 95% confidence interval. The horizontal black line represents the median.Abbreviation: BED, binge-eating disorder.

outpatient or inpatient treatment for BED at baseline (Agras et al., 2009), which likely exceeds the prevalence of specialty treatment for BED in our sample (Hudson et al., 2007). Likewise, outcomes of our naturalistic study are considerably worse than 6-year follow-up outcomes for 68 women who had DSM-IV BED at the time they received inpatient treatment for eating disorders (Fichter, Quadflieg, & Gnutzmann, 1998)

Outcomes in our naturalistic study were also generally worse than long-term outcomes for specific psychological interventions for BED in clinical trials, which is not surprising given evidence for the effectiveness of psychotherapy for BED in the long term (Hilbert et al., 2012), coupled with the finding that the majority of individuals with BED in the community do not receive treatment for their eating disorder (Coffino, Udo, & Grilo, 2019; Hay et al., 2020; Kessler et al., 2013). More specifically, after receiving cognitive-behavioral therapy or interpersonal psychotherapy for BED, a larger proportion of individuals reported sustained remission at 2 years post-treatment (Wilson, Wilfley, Agras, & Bryson, 2010) than at 2.5-year follow-up in this study, and a smaller proportion of individuals continued to binge eat at least once weekly at approximately 4 years post-treatment (Hilbert et al., 2012) than at the 5-year follow-up in this study. Less is known about the long-term effectiveness of pharmacologic treatment for BED, including lisdexamfetamine (Feltner et al., 2022; McElroy et al., 2016), the only medication that has gained regulatory approval for marketing for BED in the United States and other countries. However, in a placebo-controlled trial to assess relapse prevention after achieving response to a medication, treatment with lisdexamfetamine was associated with a markedly lower rate of relapse of BED (3.7%) than treatment with placebo (32.1%) over a six-month period (Hudson, McElroy, Ferreira-Cornwell, Radewonuk, & Gasior, 2017). Our results thus provide additional evidence that more widespread dissemination of psychotherapy and potentially efficacious pharmacologic treatment for BED would reduce the burden of disease due to BED (Moessner & Bauer, 2017), with the caveat that it is not known whether results from treatment-seeking samples generalize to the broader population.

Notably, there was no diagnostic crossover (e.g. from BED to anorexia nervosa or bulimia nervosa) in our study. This result contrasts with that of some prior prospective, community-based studies, which found that crossover from BED to bulimia nervosa was relatively common in adolescent females (Allen, Byrne, Oddy, & Crosby, 2013; Stice et al., 2013, 2009). However, another prospective, community-based study in young adult women found that crossover from BED to bulimia nervosa was relatively uncommon (Fairburn et al., 2000). Collectively, these results suggest that the likelihood of crossover from BED to bulimia nervosa may decrease with age and with longer BED duration.

In our sample, approximately one-third (35.0%) of individuals experiencing remission at 2.5-year follow-up subsequently relapsed to either full or sub-threshold BED at 5-year follow-up. Further, based on the diagnostic status transition data, time to relapse after remission was approximately 2.5 years. Taken together, these results suggest that relapse after remission is relatively common – at least among prevalent cases of BED in adults with generally higher BMI. Although methodological differences make it difficult to compare our relapse results to those of prior studies with similar follow-up periods (Fairburn et al., 2000; Stice et al., 2013, 2009), our relapse results appear similar to some (Stice et al., 2013, 2009), but higher than others (Fairburn et al., 2000).

Regarding the machine learning analyses, which aimed to build prediction models of time to remission from baseline demographic and clinical variables, no machine learning methods strongly or consistently outperformed random guessing. These results suggest that the demographic and clinical measures we included are not strongly predictive of time to remission for BED in community-based samples. These findings are consistent with a prior prospective, community-based study that found no consistent predictors of objective binge-eating episodes at both 3- and 6-month follow-up (Cachelin et al., 1999). Our findings are also consistent with machine learning results from a randomized controlled trial of behavioral weight loss *v.* stepped care in BED, where area under the receiver operator characteristic curve values from cross-validation ranged from 0.49 to 0.51 for traditional logistic regression, elastic-net, and random forest models that predicted binge-eating abstinence during the last month of treatment as a function of treatment condition, demographic information, and baseline clinical characteristics (Forrest et al., 2023). Thus, the factors that contribute to a better or worse naturalistic course of BED – or to binge-eating cessation in the treatment context – remain to be fully elucidated. One possibility is that the natural course of BED is influenced by changes in the individual's familial, social, professional, or broader environment that are not related to the individual's baseline characteristics.

Overall, our study has several strengths. To the best of our knowledge, it is the largest prospective, community-based study of individuals with BED and is more than three times larger than prior multi-year prospective, community-based studies of BED (Fairburn et al., 2000; Stice et al., 2013). Further, it includes men, who were excluded from all prior prospective, community-based studies of BED (Cachelin et al., 1999; Fairburn et al., 2000; Stice et al., 2009). Additionally, the mean age of BED onset for our participants was approximately 25 years, which is similar to the mean age of onset in large, community-based retrospective studies; in contrast, the mean age of onset was approximately 17 years in prior multi-year prospective, community-based studies (Fairburn et al., 2000; Stice et al., 2013). Likewise, 77% of participants in our study had a BMI of 30 or above at baseline, which is not far from the analogous percentage (68%) in the most recent nationally representative study (Udo & Grilo, 2018); in contrast, the analogous percentage was 22% (Fairburn et al., 2000) or possibly even lower (Stice et al., 2013) in prior community-based studies. Additionally, the collection of diagnostic status transition data allowed us to use survival methods to examine the time to remission and the time to relapse after remission. Further, a host of clinically relevant variables were measured at baseline, thereby providing numerous predictors for machine learning analysis. Also, our use of a machine learning framework – through dividing the dataset into a training and test set, and fitting multiple classes of models – allowed for more rigorous assessment of the baseline variables' predictive value, compared to simply fitting a regression model with all predictors to the entire dataset.

Our study findings should be interpreted in light of certain limitations. First, we sampled prevalent, rather than incident, cases of BED, and more transient cases of BED may have been underrepresented in our sample due to length bias in sampling. One would expect this bias to be worse for probands than relatives, given that the former were selected for inclusion in the original family study based on their BED status. However, proband status was not associated with time to remission (estimated Cox proportional hazard ratio for relative *v.* proband [95% confidence interval]: 0.97 [0.54–1.74]), nor was BED duration (for 5-year

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

increase: 0.97 [0.88–1.07]), which suggests that this source of bias may be small. Second, in our study, diagnosis was based on DSM-IV criteria, which raises the question of whether the results would generalize to BED by DSM-5 criteria, which requires only one binge-eating episode per week over three months, as opposed to two days per week with binge-eating episodes over six months in DSM-IV. Notably, within our family study sample, only three additional individuals met the broader DSM-5 criteria for BED at baseline (Hudson, Coit, Lalonde, & Pope, 2012), and including those individuals in the present follow-up study would thus have little impact on the findings as a whole. However, as noted above, using DSM-5 criteria to assess diagnostic status at follow-up timepoints would increase the percentages of individuals with (full) BED at those timepoints. Third, as implemented in many clinical drug trials, the same rater who interviewed the participant at baseline also interviewed them at the follow-up timepoints, to minimize measurement error due to intra-participant assessment variability over time, because our focus was detecting change in BED status. Thus, interviewers were not blinded to participants' prior BED status, which could have introduced bias (e.g. if interviewers were predisposed to see participants as unchanged, even if they had improved). However, the baseline assessment included extensive consensus ratings of BED and sub-threshold BED among all three interviewers, which likely helped to standardize their assessments of BED status. Fourth, the questions used to assess diagnostic changes over follow-up intervals have not been published or validated, and the accuracy of participants' reports of these changes may have been reduced by the limitations of memory over an approximately 2.5-year period. However, the same approach (i.e. focusing on changepoints along a timeline) is utilized in a highly-regarded interview assessing longitudinal changes in eating disorders (Eddy, Murray, & Thomas, 2015). Further, using only information on diagnosis status at follow-up timepoints leads to the same conclusions regarding remission and relapse in our study, with the exception of the median time to remission, which is greater than 2.5 years when using only this information. Fifth, we examined *remission* from BED based on the absence of binge-eating episodes, but we did not assess either researcher- or individually-defined *recovery*, which can reflect improvements in eating disorder cognitions and other domains instead of or in addition to cessation of regular binge eating (Bardone-Cone et al., 2010; Silén et al., 2021; Slof-Op 't Landt, Dingemans, de la Torre, Rivas, & van Furth, 2019). Sixth, importantly, although our sample was more representative of individuals with BED in the general population in terms of sex, age of onset, and BMI compared to other prospective studies (Fairburn et al., 2000; Stice et al., 2013, 2009), all but one participant in our study identified as White or Black non-Hispanic, limiting generalizability to other racial and ethnic groups (Goel et al., 2022). Likewise, only three participants in our sample had BMI less than 25 at baseline, limiting our findings' generalizability to individuals with lower BMI. Seventh, in our analyses, we considered four machine learning methods to predict time to remission, three of which belonged to the same class (i.e. Cox regression). It is possible that other methods not included in our *a priori* data analysis plan (such as support vector machines or neural networks) may have yielded better predictions, although standard Cox regression and random survival forests have demonstrated better predictive accuracy for survival outcomes than these other methods in simulations (Wang et al., 2022). Further, it is possible that predictive power may have been better with measures not included among our baseline predictors (e.g. frequency of objective binge-eating episodes, shape/weight overvaluation (Forrest, Jacobucci, & Grilo, 2022), scores on the Binge Eating Scale (Gormally, Black, Daston, & Rardin, 1982) or the Yale-Brown Obsessive Compulsive Scale modified for Binge Eating (Deal, Wirth, Gasior, Herman, & McElroy, 2015), or measures of depressive symptoms), or for outcomes other than BED remission (e.g. overall eating disorder symptoms [Forrest et al., 2023]). Additionally, our sample size, although considerably larger than existing studies, was not large in absolute terms, and less than half the sample experienced remission, limiting our ability to detect weak effects of the baseline variables and to achieve high predictive power, especially for random survival forests.

In conclusion, these limitations notwithstanding, our results suggest that, among adults with generally higher BMI in naturalistic circumstances, BED improves with time, but full remission often takes multiple years, and relapse is relatively common.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291724000977

Acknowledgments. The authors gratefully acknowledge M. Shafiqur Rahman for providing an excerpt of the Stata code used in Rahman et al. (2017).

Disclaimers. Ortho-McNeil Janssen Scientific Affairs had no control over the content of this manuscript or the authors' freedom to publish it. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

Funding statement. The family study and longitudinal study were supported in part by an investigator-initiated grant from Ortho-McNeil Janssen Scientific Affairs (JH). Further, KNJ's work on this project was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (KJ, grant number K23-DK120517) and by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. CMB's work on this project was supported in part by the National Institute of Mental Health (CB, grant numbers R56-MH129437, R01-MH120170, R01-MH119084, R01-MH118278, and R01MH124871); and the Swedish Research Council (CB, Vetenskapsrådet, grant number 538-2013-8864l).

Competing interests. Author 1 has owned equity shares in Sanofi and Centene Corporation, served on the Clinical Advisory Board for Beanbag Health, and received research funding from the National Institute of Diabetes and Digestive and Kidney Diseases. Author 8 has received consulting fees from Otsuka and has received research grant support from Boehringer-Ingelheim and Idorsia. Author 4 reports the following relationships: Lundbeckfonden (grant recipient); Pearson (author, royalty recipient). Author 6 is or has been a consultant to or member of the scientific advisory boards of F. Hoffmann-La Roche Ltd., Idorsia, Myriad, Novo Nordisk, Otsuka, Sipnose, Sunovion, and Takeda. She is or has been a principal or co-investigator on studies sponsored by Brainsway, Idorsia, Janssen, Marriott Foundation, Myriad, National Institute of Mental Health, Novo Nordisk, Otsuka, and Sunovion. She is also an inventor on United States Patent No. 6,323,236 B2, Use of Sulfamate Derivatives for Treating Impulse Control Disorders, and along with the patent's assignee, University of Cincinnati, Cincinnati, Ohio, has received payments from Johnson & Johnson, which has exclusive rights under the patent. Author 5 has received royalties from Up to Date. Author 2, Author 3, and Author 7 report no financial relationships with commercial interests.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

Agras, W. S., Crow, S., Mitchell, J. E., Halmi, K. A., & Bryson, S. (2009). A 4-year prospective study of eating disorder NOS compared with full eating disorder syndromes. *International Journal of Eating Disorders*, *42*, 565–570. doi: 10.1002/eat.20708

Allen, K. L., Byrne, S. M., Oddy, W. H., & Crosby, R. D. (2013). Early onset binge eating and purging eating disorders: Course and outcome in a population-based study of adolescents. *Journal of Abnormal Child Psychology*, *41*, 1083–1096. doi: 10.1007/s10802-013-9747-7

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

Brown, T., Klein, K., & Keel, P. K. (2015). The "natural" course of eating disorders. In L. Smolak & M. P. Levine (Eds.), *The wiley handbook of eating disorders* (1st ed., pp. 757–769). Chichester, UK: Wiley. doi: 10.1002/9781118574089.ch55

Cachelin, F. M., Striegel-Moore, R. H., Elder, K. A., Pike, K. M., Wilfley, D. E., & Fairburn, C. G. (1999). Natural course of a community sample of women with binge eating disorder. *International Journal of Eating Disorders*, *25*, 45–54. doi: 10.1002/(SICI)1098-108X(199901)25:1<45::AID-EAT6>3.0.CO;2-3

Coffino, J. A., Udo, T., & Grilo, C. M. (2019). Rates of help-seeking in US adults with lifetime DSM-5 eating disorders: Prevalence across diagnoses and differences by sex and ethnicity/race. *Mayo Clinic Proceedings*, *94*, 1415–1426. doi: 10.1016/j.mayocp.2019.02.030

Deal, L. S., Wirth, R. J., Gasior, M., Herman, B. K., & McElroy, S. L. (2015). Validation of the Yale-Brown obsessive compulsive scale modified for binge eating. *International Journal of Eating Disorders*, *48*, 994–1004. doi: 10.1002/eat.22407

Diemer, E. W., Hudson, J. I., & Javaras, K. N. (2021). More (adjustment) is not always better: How directed acyclic graphs can help researchers decide which covariates to include in models for the causal relationship between an exposure and an outcome in observational research. *Psychotherapy and Psychosomatics*, *90*, 289–298. doi: 10.1159/000517104

Eddy, K. T., Murray, H. B., Thomas, J. J. (2015). Longitudinal Interval Follow-Up Evaluation Eating and Feeding Disorders Version 3. (LIFE-EAT 3). Adapted from Keller et al., 1987.

Fairburn, C. G., Cooper, Z., Doll, H. A., Norman, P., & O'Connor, M. (2000). The natural course of bulimia nervosa and binge eating disorder in young women. *Archives of General Psychiatry*, *57*, 659–665. doi: 10.1001/archpsyc.57.7.659

Feltner, C., Peat, C., Reddy, S., Riley, S., Berkman, N., Middleton, J. C., … Jonas, D. E. (2022). Screening for eating disorders in adolescents and adults: Evidence report and systematic review for the US preventive services task force. *JAMA*, *327*, 1068–1082. doi: 10.1001/jama.2022.1807

Fichter, M. M., Quadflieg, N., & Gnutzmann, A. (1998). Binge eating disorder: Treatment outcome over a 6-year course. *Journal of Psychosomatic Research*, *44*, 385–405. doi: 10.1016/S0022-3999(97)00263-8

First, M. B., Spitzer, R. L., Gibbons, M., & Williams, J. B. W. (2001). *Structured clinical interview for axis I DSM-IV disorders*. New York: New York State Psychiatric Institute.

Forrest, L. N., Ivezaj, V., & Grilo, C. M. (2023). Machine learning *v.* Traditional regression models predicting treatment outcomes for binge-eating disorder from a randomized controlled trial. *Psychological Medicine*, *53*, 2777–2788. doi: 10.1017/S0033291721004748

Forrest, L. N., Jacobucci, R. C., & Grilo, C. M. (2022). Empirically determined severity levels for binge-eating disorder outperform existing severity classification schemes. *Psychological Medicine*, *52*, 685–695. doi: 10.1017/S0033291720002287

Garner, D. (1991). *Eating disorder inventory-2*. Odessa, FL: Psychological Assessment Resources.

Glazer, K. B., Sonneville, K. R., Micali, N., Swanson, S. A., Crosby, R., Horton, N. J., … Field, A. E. (2019). The course of eating disorders involving bingeing and purging among adolescent girls: Prevalence, stability, and transitions. *Journal of Adolescent Health*, *64*, 165–171. doi: 10.1016/j.jadohealth.2018.09.023

Goel, N. J., Mathis, K. J., Egbert, A. H., Petterway, F., Breithaupt, L., Eddy, K. T., … Graham, A. K. (2022). Accountability in promoting representation of historically marginalized racial and ethnic populations in the eating disorders field: A call to action. *The International Journal of Eating Disorders*, *55*, 463–469.

Gormally, J., Black, S., Daston, S., & Rardin, D. (1982). The assessment of binge eating severity among obese persons. *Addictive Behaviors*, *7*, 47–55. doi: 10.1016/0306-4603(82)90024-7

Hay, P., Ghabrial, B., Mannan, H., Conti, J., Gonzalez-Chica, D., Stocks, N., … Touyz, S. (2020). General practitioner and mental healthcare use in a community sample of people with diagnostic threshold symptoms of bulimia nervosa, binge-eating disorder, and other eating disorders. *International Journal of Eating Disorders*, *53*, 61–68. doi: 10.1002/eat.23174

Hilbert, A., Bishop, M. E., Stein, R. I., Tanofsky-Kraff, M., Swenson, A. K., Welch, R. R., & Wilfley, D. E. (2012). Long-term efficacy of psychological treatments for binge eating disorder. *The British Journal of Psychiatry*, *200*, 232–237. doi: 10.1192/bjp.bp.110.089664

Hudson, J. I., Coit, C. E., Lalonde, J. K., & Pope, Jr., H. G. (2012). By how much will the proposed new DSM-5 criteria increase the prevalence of binge eating disorder? *International Journal of Eating Disorders*, *45*, 139–141. doi: 10.1002/eat.20890

Hudson, J. I., Hiripi, E., Pope, Jr., H. G., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the national comorbidity survey replication. *Biological Psychiatry*, *61*, 348–358. doi:10.1016/j.biopsych.2006.03.040

Hudson, J. I., Hudson, Y., Kanayama, G., Schnabel, J., Javaras, K. N., Kaufman, M. J., … Pope, Jr., H. G. (2023). Causal factors in childhood and adolescence leading to anabolic-androgenic steroid use: A machine learning approach. *Drug and Alcohol Dependence Reports*, *10*, 100215. doi: 10.1016/j.dadr.2023.100215

Hudson, J. I., Lalonde, J. K., Berry, J. M., Pindyck, L. J., Bulik, C. M., Crow, S. J., … Pope, Jr., H. G. (2006). Binge-eating disorder as a distinct familial phenotype in obese individuals. *Archives of General Psychiatry*, *63*, 313–319. doi: 10.1001/archpsyc.63.3.313

Hudson, J. I., Lalonde, J. K., Coit, C. E., Tsuang, M. T., McElroy, S. L., Crow, S. J., … Pope, Jr., H. G. (2010). Longitudinal study of the diagnosis of components of the metabolic syndrome in individuals with binge-eating disorder. *The American Journal of Clinical Nutrition*, *91*, 1568–1573. doi: 10.3945/ajcn.2010.29203

Hudson, J. I., McElroy, S. L., Ferreira-Cornwell, M. C., Radewonuk, J., & Gasior, M. (2017). Efficacy of lisdexamfetamine in adults with moderate to severe binge-eating disorder: A randomized clinical trial. *JAMA Psychiatry*, *74*, 903–910. doi: 10.1001/jamapsychiatry.2017.1889

Javaras, K. N., Laird, N. M., Reichborn-Kjennerud, T., Bulik, C. M., Pope, Jr., H. G., & Hudson, J. I. (2008a). Familiality and heritability of binge eating disorder: Results of a case-control family study and a twin study. *International Journal of Eating Disorders*, *41*, 174–179. doi: 10.1002/eat.20484

Javaras, K. N., Pope, Jr., H. G., Lalonde, J. K., Roberts, J. L., Nillni, Y. I., Laird, N. M., … Hudson, J. I. (2008b). Co-occurrence of binge eating disorder with psychiatric and medical disorders. *The Journal of Clinical Psychiatry*, *69*, 266–273. doi: 10.4088/JCP.v69n0213

Keski-Rahkonen, A. (2021). Epidemiology of binge eating disorder: Prevalence, course, comorbidity, and risk factors. *Current Opinion in Psychiatry*, *34*, 525–531. doi: 10.1097/YCO.0000000000000750

Kessler, R., Berglund, P., Chiu, W., Deitz, A., Hudson, J., Shahly, V., … Xavier, M. (2013). The prevalence and correlates of binge eating disorder in the World Health Organization World Mental Health Surveys. *Biological Psychiatry*, *73*, 904–914. doi: 10.1016/j.biopsych.2012.11.020

Knoph, C., Von Holle, A., Zerwas, S., Torgersen, L., Tambs, K., Stoltenberg, C., … Reichborn-Kjennerud, T. (2013). Course and predictors of maternal eating disorders in the postpartum period. *International Journal of Eating Disorders*, *46*, 355–368. doi: 10.1002/eat.22088

Marshall, A., Altman, D. G., Holder, R. L., & Royston, P. (2009). Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Medical Research Methodology*, *9*, 57. doi: 10.1186/1471-2288-9-57

McElroy, S. L., Hudson, J., Ferreira-Cornwell, M. C., Radewonuk, J., Whitaker, T., & Gasior, M. (2016). Lisdexamfetamine dimesylate for adults with

https://doi.org/10.1017/S0033291724000977 Published online by Cambridge University Press

moderate to severe binge eating disorder: Results of two pivotal phase 3 randomized controlled trials. *Neuropsychopharmacology*, *41*, 1251–1260. doi: 10.1038/npp.2015.275

Moessner, M., & Bauer, S. (2017). Maximizing the public health impact of eating disorder services: A simulation study. *International Journal of Eating Disorders*, *50*, 1378–1384. doi: 10.1002/eat.22792

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*, 768–774. doi: 10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Pope, Jr., H. G., Lalonde, J. K., Pindyck, L. J., Walsh, T., Bulik, C. M., Crow, S. J., … Hudson, J. I. (2006). Binge eating disorder: A stable syndrome. *American Journal of Psychiatry*, *163*, 2181–2183. doi: 10.1176/ajp.2006.163.12.2181

Rahman, M. S., Ambler, G., Choodari-Oskooei, B., & Omar, R. Z. (2017). Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Medical Research Methodology*, *17*, Article 60. doi: 10.1186/s12874-017-0336-2

Schmid, M., Hielscher, T., Augustin, T., & Gefeller, O. (2011). A robust alternative to the Schemper-Henderson estimator of prediction error. *Biometrics*, *67*, 524–535. doi: 10.1111/j.1541-0420.2010.01459.x

Silén, Y., Sipilä, P. N., Raevuori, A., Mustelin, L., Marttunen, M., Kaprio, J., & Keski-Rahkonen, A. (2021). Detection, treatment, and course of eating disorders in Finland: A population-based study of adolescent and young adult females and males. *European Eating Disorders Review*, *29*, 720–732. doi: 10.1002/erv.2838

Slof-Op 't Landt, M. C. T., Dingemans, A. E., de la Torre, Y., Rivas, J., & van Furth, E. F. (2019). Self-assessment of eating disorder recovery: Absence of eating disorder psychopathology is not essential. *International Journal of Eating Disorders*, *52*, 956–961. doi: 10.1002/eat.23091

Stice, E., Marti, C. N., & Rohde, P. (2013). Prevalence, incidence, impairment, and course of the proposed DSM-5 eating disorder diagnoses in an 8-year prospective community study of young women. *Journal of Abnormal Psychology*, *122*, 445–457. doi: 10.1037/a0030679

Stice, E., Marti, C. N., Shaw, H., & Jaconis, M. (2009). An 8-year longitudinal study of the natural history of threshold, subthreshold, and partial eating disorders from a community sample of adolescents. *Journal of Abnormal Psychology*, *118*, 587–597. doi: 10.1037/a0016481

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research*, *29*, 71–83. doi: 10.1016/0022-3999(85)90010-8

Taylor, J. M. (2015). Psychometric analysis of the ten-item perceived stress scale. *Psychological Assessment*, *27*, 90–101. doi: 10.1037/a0038100

Taylor, J. Y., Caldwell, C. H., Baser, R. E., Faison, N., & Jackson, J. S. (2007). Prevalence of eating disorders among blacks in the national survey of American life. *International Journal of Eating Disorders*, *40*, S10–S14. doi: 10.1002/eat.20451

Udo, T., & Grilo, C. M. (2018). Prevalence and correlates of DSM-5–defined eating disorders in a nationally representative sample of U.S. adults. *Biological Psychiatry*, *84*, 345–354. doi: 10.1016/j.biopsych.2018.03.014

Wang, M., Greenberg, M., Forkert, N. D., Chekouo, T., Afriyie, G., Ismail, Z., … Sajobi, T. T. (2022). Dementia risk prediction in individuals with mild cognitive impairment: A comparison of cox regression and machine learning models. *BMC Medical Research Methodology*, *22*, Article 284. doi: 10.1186/s12874-022-01754-y

Wilfley, D. E., Wilson, G. T., & Agras, W. S. (2003). The clinical significance of binge eating disorder. *International Journal of Eating Disorders*, *34*, S96–S106. doi: 10.1002/eat.10209

Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2010). Psychological treatments of binge eating disorder. *Archives of General Psychiatry*, *67*, 94–101. doi: 10.1001/archgenpsychiatry.2009.170

Wood, A. M., Royston, P., & White, I. R. (2015). The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical Journal*, *57*, 614–632. doi: 10.1002/bimj.201400004

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. doi: 10.1177/1745691617693393