Predictors of long-term recovery in anorexia nervosa and bulimia nervosa: Data from a 22-year longitudinal study

Debra L. Franko, Nassim Tabri, Aparna Keshavia, Helen B. Murray, David B. Herzog, Jennifer J. Thomas, Kathryn Conilio, Pamela K. Keel, Kamryn T. Eddy

ABSTRACT

Objective: The objective of this study was to investigate predictors of long-term recovery from eating disorders 22 years after entry into a longitudinal study.

Method: One hundred and seventy-six of the 228 surviving participants (77.2%) were re-interviewed 20-25 years after study entry using the Longitudinal Interval Follow-up Evaluation to assess ED recovery. The sample consisted of 100 women diagnosed with anorexia nervosa (AN) and 76 with bulimia nervosa (BN) at study entry.

Results: A comorbid diagnosis of major depression at the start of the study strongly predicted having a diagnosis of AN-Restricting type at the 22-year assessment. A higher body mass index (BMI) at study intake decreased the odds of being diagnosed with AN-Binge Purge type, relative to being recovered, 22 years later. The only predictor that increased the likelihood of having a diagnosis of BN at the 22-year assessment was the length of time during the study when the diagnostic criteria for BN were met.

Conclusions: Together, these results indicate that the presence and persistence of binge eating and purging behaviors were poor prognostic indicators and that comorbidity with depression is particularly pernicious in AN. Treatment providers might pay particular attention to these issues in an effort to positively influence recovery over the long-term.

Eating disorders are a major public health concern, with significant morbidity and mortality (Smink et al., 2013). Therefore, it is important to understand the recovery process and identify factors that increase the likelihood of recovery. Recovery rates in anorexia nervosa (AN) and bulimia nervosa (BN) differ depending on recovery definition and vary widely (Herzog et al., 1992; Steinhausen, 2009). The identification of predictors of recovery could provide key targets for treatment to prevent future episodes of illness and promote the maintenance of long-term recovery.

Studies of AN have identified a number of predictors of recovery, including mood and anxiety symptoms, personality traits, body weight, and psychosocial difficulties (Keski-Rahkonen et al., 2013; Zerwas et al., 2013; Zipfel et al., 2000). Zipfel et al. (2000) examined predictors of outcome in AN (N = 84) and found that a long duration of illness prior to hospitalization, low BMI, inadequate weight gain during hospitalization, and severe psychological or social problems predicted poor outcome 21 years after hospitalization. A large scale retrospective study of 680 women examined the association between eating disorder features, personality traits, and psychiatric comorbidity in relation to recovery, as defined by one year without eating disorder symptoms (Zerwas et al., 2013). Based on survival analysis to model time to recovery, the presence of vomiting and higher trait anxiety predicted a lower likelihood of recovery. Although the study was highly powered, it was limited by its retrospective and cross-sectional design with inherent potential for recall bias. A population-based study of 55 women with AN revealed in multivariate analyses that recovery was less likely when depressive symptoms were present prior to the onset of the eating disorder; however, other predictors (unemployment, marital dissatisfaction, perfectionism) were no longer significant once duration of eating disorder illness was controlled for in the analyses (Keski-Rahkonen et al., 2013). Steinhausen (2009) noted an association between positive family relationships and outcome in AN, finding that...
negative family ties were linked prospectively to the onset of binge eating and vomiting, suggesting that the relationship between vomiting and poorer outcome may be confounded by family-related and within-subject emotional factors.

For BN, research on predictors is inconsistent and lacks a long-term perspective. In a study by Lock et al. (2013), recovery for adults with BN was best predicted by a low frequency of compensatory behaviors (fewer than 2 times/month), whereas for adolescents, purging abstinence and lower Eating Disorder Examination restraint scores (EDE; Fairburn and Beglin, 1994) predicted recovery from BN. Keel and colleagues identified longer duration of illness and a history of substance use problems as predictors of poor outcome in a study of BN 11.5 years after diagnosis (Keel et al., 1999). A comprehensive review of definitions of treatment outcome in BN identified eight studies examining predictors of outcome (Williams et al., 2012). Three studies examined predictors of outcome for cognitive-behavioral treatment for BN, finding that early progress (significant decrease in vomiting, Agras et al., 2000; percent reduction in binge eating at 2, 4, 6, and 8 weeks of treatment, Marrone et al., 2009), weight suppression (Butryn et al., 2006; Lowe et al., 2011; but not by Carter et al., 2008), and attitudes toward shape and weight and self-esteem (Fairburn et al., 1993) predicted positive outcome. Rowe et al. (2011) found that for 109 of 134 participants in a psychotherapy treatment trial, self-directedness was the only predictor of any eating disorder diagnosis (past year) at 5-year follow-up, such that high self-directedness predicted a greater likelihood of recovery. Recovery in family-based treatment for adolescent BN was predicted by lower EDE eating concerns scores at baseline (Le Grange et al., 2008). Thus the literature on predictors of recovery in BN is characterized by varied findings across studies, which is likely due to differences in sample characteristics and length of follow-up, as well as varying definitions of recovery.

Comorbidity may decrease the likelihood of recovery (e.g., Keski-Rahkonen et al., 2013). Eating disorders (EDs) have high rates of comorbidity, particularly with depression and substance abuse. Major depressive disorder (MDD) is the most common comorbid diagnosis in patients with EDs (Berkman et al., 2007; Kaye, 2008; Löwe et al., 2001; Puccio et al., 2016; Yager et al., 2005). Additionally, high rates of substance use disorder (SUD) have been found to occur among women with eating disorders, including AN, BN, and binge eating disorder (BED; Munn-Chernoff and Baker, 2015). For instance, in one of the largest epidemiological explorations of this association using data from the National Comorbidity Survey Replication, Hudson and colleagues (2007) found the prevalence of SUDs to be as high as 27%, 37% and 23% in individuals with AN, BN and BED, respectively. Individuals with comorbid EDs and SUDs are susceptible to increased morbidity and mortality, with alcohol abuse increasing the risk for mortality over twofold in AN (Franko et al., 2013), for example. Given high rates of comorbidity in eating disorder samples, and resultant potential long-term negative consequences of such, we were interested in the extent to which depressive and substance use disorders might predict non-recovery. Although other comorbidities, such as anxiety disorders and medical issues are important and common (Keski-Rahkonen and Mustelin, 2016), only depression and substance abuse were measured at intake and throughout the study and thus are the only comorbid disorders that could be examined as potential predictors of recovery in this study.

Overall, the existing literature on the predictors of recovery is characterized by variability in sample characteristics and sample size, disparate operational definitions and measures of recovery, and varied length of follow-up. Thus, the purpose of this study was to examine predictors of recovery in AN and BN in a large longitudinal sample over an extended period of time, which allowed for a sensitive assessment of course of illness and time to recovery. Our aim in this report was to identify intake and course predictors of recovery at a one-time 22-year assessment.

1. Method

1.1. Participants

Participants were recruited from Boston-area outpatient eating disorder services (1987-1991) if they met the following inclusion criteria: DSM-III-R (American Psychiatric Association, 1987) AN or BN diagnosis; female; minimum age of 12 years; residence within 200 miles of Boston; English speaking; and no evidence of organic brain syndrome or terminal illness.

Of the 294 women meeting study criteria, 250 (85.0%) agreed to participate, and four dropped out prior to the first follow-up. Thus, the study group comprised 246 women. Retrospectively applying DSM-IV (American Psychiatric Association, 2000) criteria to intake data, 51 met criteria for AN restricting type (AN-R), 85 met criteria for AN binge/purge type (AN-BP), and 110 met criteria for BN. At baseline, the mean age of the sample was 24.8 years (range 13–45 years) and the mean duration of illness was 6.7 years (range 3 months–21 years). Average age of onset was 18 years. The sample was all female and predominantly white (96%). Within the first five years of the study, only 11 participants (4%) reported not receiving some form of treatment (inpatient or outpatient). Women with AN spent significantly more time in inpatient and group treatment than did women with BN (Keel et al., 2002).

Participants were re-contacted between 2011 and 2013, which was 20–25 years after study intake. Of the initial sample of 246, 18 had died by the Wave II assessment. Of these 18, 15 had been diagnosed with AN at intake and 4 deaths were due to suicide. For the remaining 14, three died due to alcohol or drug overdose, four deaths were related to cardiac issues, five were related to a variety of medical issues (amyotrophic lateral sclerosis, gastrointestinal hemorrhage, brain injury, pneumonia, cerebral glioma), and we were not able to ascertain cause of death in two cases. Predictors of mortality included alcohol abuse, low BMI, and poor social adjustment (Franko et al., 2013).

Of the 228 survivors, 176 (77%) participated at 25-year follow-up, 37 (16%) were contacted but declined participation, and 15 (7%) were lost to follow-up. Of these 176, 40 had a diagnosis of AN-R, 60 had a diagnosis of AN-BP, and 76 had a diagnosis of BN at study intake. Compared to the 21 women with a diagnosis at intake of AN who did not participate in Wave II, the 100 participants with AN had a shorter duration of illness at intake (5 v. 8 years, p < 0.03) but did not differ in baseline BMI or age. Compared to the 31 women with BN who didn’t participate in Wave II, the 76 participants had a lower intake BMI (22 v. 25, p < 0.02), younger age of onset (17 v. 19 years, p < 0.03), and higher rates of recovery at the end of Wave I (p < 02); however, they did not differ in age.

1.2. Study procedures and measures

During Wave I of data collection (1987-1996), participants were interviewed every 6-12 months for a mean of 9.1 (SD = 1.6) years. In Wave II of data collection (2011-2013), the surviving participants were re-contacted between 20 and 25 years after study initiation. For the 176 participants, the mean length of follow-up since intake was 22.1 (SD = 1.1) years.

All participants completed the Longitudinal Interval Follow-Up Evaluation Eating Disorders Version (LIFE-EAT II; Herzog et al., 1999). The LIFE-EAT II is a semi-structured interview based on the LIFE (Keller et al., 1987) and was administered by trained research coordinators under the supervision of study psychologists and psychiatrists at intake and every 6-12 months during Wave I (1987-1996) and again one time at Wave II (2011-2013). At each follow-up interview, the assessor determined whether participants had eating disorder symptoms and behaviors, and screened for major areas of psychopathology during the current week and during each of the preceding weeks since the last interview based on participants’ retrospective recall, anchored by use of
key dates and intervening life events. Specifically, participants were asked about their eating disorder symptoms each week for the time since the last interview (Wave I) and for each week during the previous year (Wave II), using specific anchors, life events, holidays, and key dates (e.g., returning to school in September). During Wave I, interviews were administered in-person whenever possible. In Wave II, all interviews were administered by telephone to assess symptoms in the past year.

Based on this retrospective recall, the LIFE-EAT II yielded weekly psychiatic status rating (PSR) scores. PSR scores are ordinal, symptom-oriented scale scores based on Research Diagnostic Criteria (Spitzer et al., 1978) ratings for AN and BN that ranged from 1 (complete recovery) to 6 (definitely meets diagnostic criteria—severe). Weekly AN and BN PSR scores are derived by the interviewer, based on the pattern of responses by the participant to the individual symptom questions in the LIFE-EAT II. A PSR score of 1 or 2 indicates few or no symptoms (e.g., 95% expected body weight for AN, no binge or purge behaviors for BN), a PSR score of 3 or 4 indicates significant symptoms and narrowly missing the diagnostic criteria, and a PSR score of 5 or 6 indicates meeting full criteria for AN or BN, with 6 indicating symptoms severe enough to warrant hospitalization.

Height and weight were measured by research staff only at intake. Correlations between measured height and weight and self-reported height and weight were significant (p < 0.001), supporting the validity of using self-reported weights throughout the study (Murray et al., 2017). For the present study, we examined all of the available Wave I intake, course, and last visit variables from the LIFE-EAT, as listed in each section below, as predictors of rates of recovery at Wave II.

**Intake variables.** We included participants’ age in years, body mass index (BMI), and eating disorder diagnosis (AN-R, AN-BP, and BN) as well as whether or not participants had a comorbid diagnosis of major depressive disorder (MDD), alcohol use disorder, and drug use disorder (Spitzer et al., 1978).

**Course variables.** We included percentage of weeks during which participants met diagnostic criteria for AN and BN, respectively, during Wave I. We also included the percentage of weeks during which binge eating and purging behaviors occurred during Wave I.

In addition to eating disorder symptoms and behaviors, we included the percentage of weeks participants met criteria for major depressive disorder (MDD) and substance use disorder. Only depression and substance use were assessed in this study both at intake and throughout Wave I, in order to decrease subject burden which may have occurred if the full range of comorbid diagnoses had been assessed at each visit. MDD was assessed during the LIFE-EAT II interview to attain PSR scores that ranged from 1 (usual self) to 6 (definitely meets diagnostic criteria—severe). We calculated the percentage of weeks during which participants had a PSR score value ≥ 5, which corresponds to definitely meets diagnostic criteria on the PSR response scale. Moreover, we included the percentage of weeks during which participants met criteria for alcohol use disorder and drug use disorder, respectively. Both these diagnoses were assessed using PSR that ranged from 1 (not present) to 3 (definitely meets diagnostic criteria—severe) and we calculated the percentage of weeks during which participants had a PSR score of 3.

To examine psychosocial functioning as a predictor of recovery, we included the average global social adjustment score across Wave I for each participant based on the Longitudinal Interval Follow-up Evaluation-Range of Impaired Functioning Tool (LIFE-RIFT; Leon et al., 1999), which was used to assess functional impairment. This measure is interview-based and assesses level of impairment in multiple domains (e.g., work, interpersonal relations, recreation) and determines a global social adjustment score that ranges from 1 (no impairment) to 5 (severe impairment).

Importantly, we included the average number of outpatient treatment sessions per month over the course of Wave I to examine the role of treatment as a predictor of recovery.

**Last visit variables.** We included participants’ PSR scores for MDD, alcohol use disorder, and drug use disorder from the last time they were interviewed during Wave I. We also included participants’ global social adjustment score at this last visit.

**Wave II recovery status.** Long-term recovery was defined as a PSR ≤ 2 for both Anorexia and Bulimia PSR scores for the duration of the 52-week data collection period (Eddy et al., 2017). Non-recovery status for AN-R was defined by having an AN PSR ≥ 3 and BN PSR ≤ 2 at the end of the data collection period. Non-recovery status for AN-BP was defined by having an AN PSR ≥ 3 and BN PSR ≥ 3 at the end of the data collection period. Non-recovery status for BN was defined by having an AN PSR ≤ 2 and BN PSR ≥ 3 at the end of the data collection period.

Our definition of recovery (maximum PSR = 1, the absence of symptoms or PSR = 2, the presence of only mild residual symptoms over 52 weeks) has shown concurrent validity in our recent report of recovery rates in this sample (Eddy et al., 2017). We demonstrated that recovered participants at Wave II had lower EDE-Q scores and higher quality of life scores than non-recovered participants in the sample. Moreover, our definition of recovery is consistent with and in some cases more conservative than others who have examined and validated recovery definitions for both AN and BN (Bardone-Cone et al., 2010; Keel et al., 2000; Williams et al., 2012). Finally, the DSM-5 (American Psychiatric Association, 2013) defines full remission as when none of the diagnostic criteria have been met for a sustained period of time. We believe that a period of one year with no or few symptoms that do not meet the diagnostic criteria satisfies this definition.

**1.3. Data analysis**

Our goal was to identify the intake, course, and last visit predictors of long-term recovery at Wave II. A multinomial regression analysis was conducted across all participants to examine the intake, course, and last visit predictors of non-recovered status (AN-R, AN-BP, and BN) relative to recovery at Wave II. In this analysis, the reference category was recovered at Wave II and we examined predictors of having AN-R, AN-BP, or BN relative to being recovered at Wave II. Intake diagnosis was also included as a predictor and the intake, course, and last visit predictors were entered simultaneously in the regression. This procedure allowed us to examine the unique relationship between each predictor and the odds of having AN-R, AN-BP, or BN relative to being recovered at Wave II, while statistically controlling for shared variance between the predictors. Analyses indicated that intake diagnosis (AN-R, AN-BP, and BN) did not moderate the link between each predictor and recovery status at Wave II.

There were missing data for recovery status at Wave II. Specifically, of the 228 participants, only 176 participated in Wave II (77% of the original cohort). Thus, we had no data for 52 participants (22.8%). As well, one participant’s BMI at intake was missing. To include participants with missing data in the analyses, multiple imputation procedures in SPSS (v. 21) were used. In line with recommendations by Enders (2011) and Graham, (2009), we imputed missing values for recovery status at Wave II (i.e., AN-R, AN-BP, BN, or recovered) and BMI at intake using 20 data sets to enhance power. As well, intake, course, and last visit predictors were included in the imputation model to enhance the estimation of the missing data.

All estimates from multinomial logistic regression are based on a pooling of estimates across 20 data sets. The effect size calculated for the predictors was the odds ratio (OR) because the dependent variable was nominal. All analyses were conducted using SPSS (v. 21).

**2. Results**

**2.1. Recovery rates at 22 years**

Recovery status at Wave II for the observed and imputed data is presented in Table 1. At the 22-year follow-up, nearly two-thirds of the
Table 1
Recovery status at 22-years for observed and imputed data.

<table>
<thead>
<tr>
<th></th>
<th>Recovered</th>
<th>AN-R</th>
<th>AN-BP</th>
<th>BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed (N = 176)</td>
<td>113</td>
<td>23</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>Imputed (N = 228)</td>
<td>135.5</td>
<td>28.8</td>
<td>20.2</td>
<td>43.6</td>
</tr>
</tbody>
</table>

Note. AN-R = anorexia nervosa-restricting type; AN-BP = anorexia nervosa-binging and purging type; BN = bulimia nervosa. Imputed descriptives are based on 20 imputed data sets.

Table 2
Results of a multinomial logistic regression examining Wave I predictors of non-recovery status (AN-R, AN-BP, and BN) relative to recovery at Wave II.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Recovered vs. AN-R</th>
<th>Recovered vs. AN-BP</th>
<th>Recovered vs. BN</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>B</td>
<td>p</td>
<td>OR</td>
</tr>
<tr>
<td>Intake variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.04</td>
<td>0.04</td>
<td>0.36</td>
</tr>
<tr>
<td>BMI</td>
<td>1.01</td>
<td>0.01</td>
<td>0.85</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN-R vs. BN</td>
<td>14.75</td>
<td>2.69</td>
<td>0.02</td>
</tr>
<tr>
<td>AN-BP vs. BN</td>
<td>3.26</td>
<td>1.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Comorbid MDD</td>
<td>5.87</td>
<td>1.77</td>
<td>0.01</td>
</tr>
<tr>
<td>Comorbid alcohol abuse</td>
<td>1.52</td>
<td>0.42</td>
<td>0.66</td>
</tr>
<tr>
<td>Comorbid drug abuse</td>
<td>1.53</td>
<td>0.43</td>
<td>0.73</td>
</tr>
<tr>
<td>Wave I course variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of weeks AN PSR ≥ 5</td>
<td>1.01</td>
<td>0.01</td>
<td>0.30</td>
</tr>
<tr>
<td>% of weeks BN PSR ≥ 5</td>
<td>1.01</td>
<td>0.01</td>
<td>0.36</td>
</tr>
<tr>
<td>% of weeks MDD PSR ≥ 5</td>
<td>0.98</td>
<td>-0.01</td>
<td>0.62</td>
</tr>
<tr>
<td>% of weeks alcohol abuse PSR ≥ 5</td>
<td>1.03</td>
<td>0.03</td>
<td>0.68</td>
</tr>
<tr>
<td>% of weeks drug abuse PSR ≥ 3</td>
<td>0.79</td>
<td>-0.22</td>
<td>0.51</td>
</tr>
<tr>
<td>Av. global social adjustment</td>
<td>1.63</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Av. therapy sessions per month</td>
<td>0.99</td>
<td>-0.01</td>
<td>0.91</td>
</tr>
<tr>
<td>Wave I last visit variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD PSR</td>
<td>0.86</td>
<td>-0.14</td>
<td>0.55</td>
</tr>
<tr>
<td>Alcohol abuse PSR</td>
<td>2.20</td>
<td>0.79</td>
<td>0.36</td>
</tr>
<tr>
<td>Drug abuse PSR</td>
<td>5.77</td>
<td>1.75</td>
<td>0.32</td>
</tr>
<tr>
<td>Global social adjustment</td>
<td>2.04</td>
<td>0.71</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Note. AN-R = anorexia nervosa-restricting type; AN-BP = anorexia nervosa-binging and purging type; BN = bulimia nervosa; OR = odds ratio; MDD = major depressive disorder.
disorders, but this is the first study to find that depression predicts an active eating disorder over such a long period of time. It is possible that the combination of depression and AN makes an individual much less likely to be able to benefit from treatment for an eating disorder, or that the interaction of the two disorders makes it very difficult to recover. To date, treatment approaches that address both depressive and AN symptoms simultaneously are yet to be developed; our data would suggest that this may be a promising avenue for research into ways to promote long-term recovery and that clinicians might need to be focused on ensuring that both illnesses receive treatment. Given that depression at intake was a unique predictor of a chronic course in AN-R and that this was not true for intake diagnoses of AN-BP or BN, there may be etiologic and nosologic implications of these findings as well (e.g., a common genetic pathway). When depression and AN co-occur clinicians need to pay particular attention to this critical combination of pathologies in striving for positive long-term outcome.

Study strengths include the high retention rate (77%), lengthy study period, the use of structured clinical interviews and the examination of psychosocial functioning and comorbidity (Ackard et al., 2014; Jenkins et al., 2014). Both the long duration of follow-up and relatively low rate of attrition increased the probability that events related to recovery were captured.

However, there are a number of limitations to our study, not the least of which was the lack of information on anxiety disorders and other psychiatric and medical comorbidities. An important limitation was our inability to reach some participants, which restricts generalizability. Generalizability is further limited due to the fact that we studied treatment-seeking individuals who began the study in 1987, potentially resulting in cohort effects. And because the participants were from the same region and seeking tertiary care at a major medical center in an urban area, our findings cannot be applied to all groups. We were not able to conduct the interviews in person at the Wave II assessment, which may have affected the data collected. Moreover, because the Wave II data were collected only for the previous year, the clinical state of participants represents only a cross-section period of time; we don’t know what percentage were chronically and persistently symptomatic over the period between the end of Wave I and the Wave II assessment. Our interviews required participants to recall their symptoms (e.g., over the past 6-12 months), which created the potential for recall bias and may have resulted in data that were limited based on the ability to remember specific symptoms in the given time frame. This is a major limitation to the study. Conclusions from this study should be considered tentative in light of these limitations.

In summary, predictors of long-term recovery in eating disorders indicate that the presence and persistence of binge eating and purging behaviors are poor prognostic indicators. For AN, comorbidity with depression is particularly pernicious. Such findings suggest that treatment providers might pay particular attention to these issues in an effort to positively influence recovery over the long-term.

Financial support

Support for this study came from grants from the National Institute of Mental Health (D.B.H., R01 MH-38333): and (D.B.H. and K.T.E., R03 MH-094832).

Conflicts of interest

None.

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.

Acknowledgements

The authors wish to thank the participants for giving their time to help us learn more about the course and outcome of eating disorders and the research assistants who have worked on this project throughout the years.

References


Carter, F.A., McIntosh, V.V., Joyce, P.R., Bulik, C.M., 2008. Weight suppression predicts weight gain over treatment but not treatment completion or outcome in bulimia nervosa. J. Abnorm. Psychol. 117 (4), 936–940. https://doi.org/10.1037/a0013942.


