

Long-term Outcome of Bulimia Nervosa

Pamela K. Keel, PhD; James E. Mitchell, MD; Kathryn B. Miller, MA; Traci L. Davis; Scott J. Crow, MD

Background: Since bulimia nervosa's introduction to the psychiatric nomenclature in 1979, data concerning long-term outcome have been largely unavailable.

Methods: Women with the diagnosis of bulimia nervosa between 1981 and 1987 who participated in 1 of 2 studies were located and invited to participate in follow-up assessments.

Results: More than 80% of the women from these studies participated in follow-up assessments and the results represent findings for 173 women. More than 10 years following presentation (mean \pm SD length of follow-up, 11.5 \pm 1.9 years), 11% of this sample met full criteria for bulimia nervosa, and 0.6% met full criteria for anorexia nervosa. An additional 18.5% met criteria for eating disorder not otherwise specified, and 69.9% of this

sample were either in full or in partial remission. For predictive factors, only the duration of the disorder at presentation and history of substance use problems demonstrated prognostic significance. Baseline treatment condition was not associated with remission of disordered eating symptoms by the follow-up assessment.

Conclusions: The findings suggest that the number of women who continue to meet full criteria for bulimia nervosa declines as the duration of follow-up increases; approximately 30%, however, continued to engage in recurrent binge eating or purging behaviors (incidence rate, 0.026 cases per person-years). A history of substance use problems and a longer duration of the disorder at presentation predicted worse outcome.

Arch Gen Psychiatry. 1999;56:63-69

THE TERM *bulimia nervosa* was introduced into the medical literature by Russell¹ in 1979. Although the short-term treatment outcome of individuals with bulimia nervosa has been well studied, data concerning long-term outcome have been largely unavailable. Indeed, the most recent edition of the *DSM-IV*² stated that, "The long-term outcome of bulimia nervosa is not known." These limited data leave many questions unanswered, including the following: what percentage of women recover from this disorder, maintain a partial syndrome, or continue to suffer from the full syndrome more than a decade after diagnosis; what characteristics of the disorder and the individual predict long-term outcome; and which, if any, treatment approaches achieve superior long-term outcomes?

A recent review³ of outcome studies suggested that from 5 to 10 years following presentation, approximately 50% of women recovered from their disorder, while almost 20% continued to meet full

criteria for bulimia nervosa. The effect of treatment has varied across follow-up studies, with some articles⁴ indicating an association between baseline treatment conditions and eating disorder outcome and others⁵ finding no long-term effect of treatment. Few prognostic factors have been replicated across studies.³

The conclusions concerning long-term outcome in the review³ were strongly limited by the existence of only 1 study of 10 years' duration describing outcome for 44 women with the diagnosis of bulimia nervosa.⁵ Our objectives in this study were to describe and find predictive factors of long-term outcome of women with bulimia nervosa.

RESULTS

MORTALITY

Of the 200 women for whom status was ascertained at follow-up, 1 woman was dead (0.5%). The cause of death was suicide by overdose and occurred within 2 years of baseline assessment. Social Secu-

From the Department of Psychology, Harvard University, Cambridge, Mass (Dr Keel); University of North Dakota and Neuropsychiatric Research Institute, Fargo (Dr Mitchell); and the Departments of Psychiatry and Psychology, University of Minnesota, Minneapolis (Ms Miller and Davis and Dr Crow).

PARTICIPANTS AND METHODS

PARTICIPANTS

Participants (N=222) from 2 previous studies^{6,7} were sought for participation in this study. Subjects were initially evaluated at the University of Minnesota's Eating Disorders Clinic, Minneapolis, between 1981 and 1987. At baseline, all subjects were required to meet DSM-III⁸ criteria for bulimia with an additional criterion of binge eating coupled with purging at least 3 times per week during the 6 months preceding evaluation. The first study⁶ assessed women from 2 to 5 years following presentation. The second investigation⁷ was a controlled treatment study. Additional inclusion and exclusion criteria for these subjects are published in the original articles.^{6,7}

Of the 222 subjects sought for participation, 1 (0.5%) was dead, 1 (0.5%) was severely disabled, 22 (9.9%) could not be located, 21 (9.5%) declined or did not complete participation, and 177 (79.7%) participated. The ascertainment rate was 90.1%. The participation rate for this study (excluding subjects who were dead or severely disabled) was 80.5%. No significant differences in participation rates existed between the subsamples^{6,7} (χ^2_1 , 0.49; $P=.49$) or baseline treatment conditions⁷ (χ^2_3 , 3.50; $P=.32$). Finally, subjects who participated did not differ from subjects who did not participate on any of the following demographic variables assessed at baseline: age (t_{220} , 0.02; $P=.99$); race (χ^2_3 , 1.60; $P=.66$); marital status (χ^2_3 , 0.66; $P=.88$); occupational level (χ^2_3 , 2.30; $P=.51$); and educational level (χ^2_2 , 1.44; $P=.49$). Of the 177 participants, 4 had never met full DSM-IV criteria for bulimia nervosa based on initial assessments and follow-up structured clinical interviews for DSM-IV (SCID-I) and were eliminated from analyses, resulting in a sample size of 173 subjects.

There were no significant differences across demographic variables between the 2 subsamples ($P>.10$) with the exception of age at follow-up assessment (t_{160} , 2.88; $P=.005$). Age at follow-up was reanalyzed with duration of follow-up as a covariate, and no significant difference remained ($F_{1,159}$, 3.11; $P=.08$). Mean (\pm SD) age for the total sample was 35.3 (5.1) years. The sample was predominantly white (99%) with only 2 nonwhite participants. All but 1 participant had graduated from high school, 30% had completed college, and 15% had completed graduate school. Most of the sample described their occupational level as administrative (33.5%) or clerical/

sales (26.6%), with fewer than 10% of each group reporting work in manual jobs or major professional positions. Approximately 75% had married, and 50% of all subjects were still in their first marriage.

PROCEDURE

Participants were initially contacted by letters from 1 of us (J.E.M.). Letters asked subjects to complete questionnaires at home and participate in a personal interview conducted at the Eating Disorders Research Office, their home, or by telephone. During personal interviews, weight was measured using a digital scale and height was measured using a physician's standard scale for office interviews and a tape measure for home interviews. Participants removed their shoes and coats but otherwise remained fully clothed during height and weight measurements. Personal interviews were conducted with 54% of participants and telephone interviews were conducted with 46% of participants. Among individuals selecting telephone interviews, 74% lived either out of state or more than 3 hours from the research office. Analyses of eating disorder diagnostic status at follow-up determined from the interviews revealed no significant difference between personal and telephone interviews (χ^2_2 , 2.37; $P=.31$). In addition, no significant differences were found between interview types and interview assessment of depression or lifetime diagnoses of affective, anxiety, substance use, or impulse control disorders ($P=.07-.90$). Written consent from each subject was obtained prior to the interview, at the time questionnaires were received. Interviews were audiotaped to determine reliability.

MEASURES

The following measures were available from initial evaluations^{6,7}: marital status, social class ranking, weight, height, Eating Disorders Questionnaire responses,⁹ and 24-item Hamilton Depression Rating Scale scores.¹⁰ The Hamilton Anxiety Rating Scale¹¹ was available from initial evaluations of the 125 women participating in the treatment outcome study.⁷

At follow-up assessment, the Eating Disorders Questionnaire was readministered, participants were reassessed on the Hamilton Depression Rating Scale, height and weight measurements were taken again, and the following tests were administered: Multidimensional Personality Questionnaire Scale 8: Control/Impulsiveness,¹² Body Shape

ity Number Death Master File and Minnesota Death Certificate searches revealed no other deaths among women lost to follow-up. A second woman suffered a ruptured cerebral aneurysm in 1990 that left her blind and severely disabled. She had required residence in a nursing home since the event. Both women were originally part of the 2- to 5-year follow-up sample.⁶

EATING DISORDER OUTCOME

There was no significant difference between the 2 subsamples for eating disorder outcome according to either the narrow (χ^2_3 , 2.95; $P=.40$) or broad (χ^2_3 , 1.90; $P=.59$) definitions of outcome. The source of sample was not significantly asso-

ciated with the number of months since last eating disorder symptoms ($F_{1,170}$, 0.04; $P=.85$) after controlling for follow-up duration. Thus, results are presented for the total sample.

At follow-up, 1 woman (0.6%) met full criteria for anorexia nervosa and 19 women (11%) met full criteria for bulimia nervosa. Thirty-two women (18.5%) met criteria for an eating disorder not otherwise specified, including 1 woman who met full criteria for binge-eating disorder. According to the narrow definition of remission, 49 women (28.3%) were in partial remission at follow-up, and 72 women (41.6%) were in full remission. According to the broad definition of remission, 40 women (23.1%) were in partial remission, and 81 (46.8%) were in full remission.

Questionnaire (BSQ),¹³ and the DSM-IV SCID-I/P,¹⁴ for Axis I disorders, with an addendum assessing impulse control disorders.

All interviews were conducted privately by 1 of us (P.K.K. or K.B.M.) or a trained research assistant. Interviewers were trained in conducting SCID-I interviews using the SCID Training Tapes prepared by New York State Psychiatric Institute for DSM-IV, and supervised by a licensed clinical psychologist. Reliability estimates are presented in **Table 1**.

DATA MANAGEMENT AND ANALYSES

For measures involving scales, responses to at least 80% of items were required for inclusion of these data in analyses. Therefore, sample sizes vary across statistical analyses of different measures. When at least 80% but fewer than 100% of items were present, scale scores were prorated. Parametric and nonparametric tests were used to assess the significance of differences in means and proportions. Thresholds for statistical significance were set at α less than .01 due to the large number of analyses conducted. In addition, the family-wise error rate for type I errors was controlled with Dunn test corrections. Data were analyzed using the Statistical Package for the Social Sciences (SPSS Inc, Chicago, Ill) for Macintosh.

INDICES OF EATING DISORDER OUTCOME

Eating disorder outcome was defined with both a narrow and broad definition of full remission. The narrow definition of full remission required freedom from disordered eating behaviors for at least 6 months; additionally, weight and shape could not unduly influence how the subject felt about or evaluated herself. During SCID administration, subjects ranked aspects that influenced how they evaluated themselves. If weight and shape were cited as main aspects of self-evaluation, this criterion was coded as present. Comparisons of women in full and partial remission according to this definition revealed a significant difference in BSQ scores (t_{109} , 4.54; $P < .001$). The broad definition of full remission required the absence of disordered eating behaviors for at least 8 weeks with no restrictions based on influence of weight or shape on self-evaluation. Women who did not meet criteria for full remission and did not meet DSM-IV criteria for an eating disorder (eating disorder not otherwise specified, bulimia nervosa, or

anorexia nervosa) were considered to be in partial remission. The minimum frequency of disordered eating behaviors for a diagnosis of eating disorder not otherwise specified was an average of once per month for 3 months. For analyses of prognostic factors, eating disorder outcome was redefined as a dichotomous variable: remission (including partial and full remission) and eating disordered (DSM-IV criteria for eating disorder diagnoses). Demographic variables did not differ significantly between women in partial or full remission ($P > .10$) with 1 exception. Under the broad definition of full remission, age at follow-up differed significantly between women in partial and full remission (t_{110} , 3.08; $P = .003$). Computing a dichotomous variable for outcome removed distinctions between narrow and broad definitions of full remission and minimized the family-wise error rate by deciding a priori the most important distinction among the 4 groups was between women who engaged in recurrent disordered eating behaviors and women who did not engage in these behaviors. Eating disorder outcome was also investigated as a continuous variable representing the log of months since the last binge eating or purging episode. (Notably, age and months since last disordered eating symptom were not significantly associated: r , -0.08; n , 162; $P = .27$).

DURATION OF FOLLOW-UP PERIOD

Mean (\pm SD) duration of follow-up for the total sample was 11.5 (\pm 1.9) years. Follow-up duration differed significantly between the 2 subsamples (t_{121} , 27.70; $P < .001$, $P < .001$). The test for heterogeneity of variance was significant, resulting in a decrease in the degrees of freedom and the value of t . Also, the second P value indicates the Dunn test correction to control for family-wise error rate. The influence of follow-up duration on eating disorder outcome was assessed for both categorical and continuous measures. The duration of follow-up was not significantly associated with the categorical measure of eating disorder outcome (mean \pm SD follow-up duration in years for women in remission, 11.66 [\pm 1.91] and eating disordered women, 11.12 [\pm 1.91]; t_{171} , 1.73; $P = .09$) but was significantly associated with the continuous measure of eating disorder outcome (r , 0.21; n , 173; $P = .005$). For prediction of continuous outcome, test statistics are reported for categorical (F , df , and P) and continuous (ΔR^2 , n , and P) prognostic variables after controlling for variance explained by the duration of follow-up.

WEIGHT AND BODY IMAGE

No significant differences in body mass index (BMI or Quetelet index), calculated as weight in kilograms divided by the square of the height in meters, existed between the 2 subsamples at baseline or follow-up assessment ($P > .10$); however, significant differences in weight variables did appear between baseline and follow-up assessments for the total sample (**Table 2**). Despite being statistically significant, the 2.5-kg change in weight is unlikely to be clinically significant because weight for most women increases as they approach middle age. Eating disorder outcome measured categorically or continuously was not significantly associated with any weight vari-

able (with 1 exception all were $P > .10$). At a trend level, women with eating disorders desired a lower weight compared with women in remission (t_{148} , -2.44; $P = .02$).

There was no significant difference between the 2 subsamples for BSQ scores (t_{159} , 0.62; $P = .54$). For the total group, the mean (\pm SD) BSQ score was 86.8 (36.7). This value is not significantly different from the mean BSQ score reported for a community sample of 535 women (mean, 81.5 [28.4]; t_{694} , 1.7; $P = .09$; $P > .10$) but is significantly lower than the mean score for a sample of 38 women with bulimia nervosa (mean, 136.9 [22.5]; t_{197} , 10.8; $P < .001$; $P < .001$).¹³ Notably, the mean age of the follow-up sample differed from that of the community and bulimic samples; however, no significant relation-

Table 1. Reliability of Measures

Measures	Reliability Estimates		
	Baseline <i>r</i> (N)	Follow-up <i>r</i> (N)	Lifetime κ
Scale scores (Cronbach α)			
Hamilton Depression Rating Scale	0.84 (170)	0.88 (173)	...
Hamilton Anxiety Rating Scale	0.83 (100)
Multi-Dimensional Personality Questionnaire	...	0.90 (155)	...
Body Shape Questionnaire	...	0.97 (161)	...
SCID diagnostic categories*			
Eating disorders†	0.89
Mood disorders	0.91
Substance use disorders	0.97
Anxiety disorders	0.91
Impulse control disorders	1.00
All diagnoses	0.90

*Reliability estimates based on interviews from 17 randomly selected subjects (10% of sample). SCID indicates structured clinical interviews for DSM-IV.

†Current eating disorder diagnoses, $\kappa = 0.88$.

ship between age and BSQ scores was found in the current sample ($r, 0.0023$; $n, 161$; $P=.98$) and no association between BSQ scores and age was reported for the comparison samples.¹³ Body Shape Questionnaire scores were significantly associated with eating disorder outcome measured both categorically ($t_{159}, 7.82$; $P<.001$; $P<.001$) and continuously ($r, -0.47$; $n, 161$; $P<.001$; $P<.001$) with symptomatic women expressing greater body shape concerns than women in remission.

PROGNOSTIC FACTORS

Age of Onset, Duration, and Severity of Symptoms

Age of onset for the disorder was ascertained from the baseline Eating Disorders Questionnaire. Correcting for outliers, the mean (\pm SD) age of onset for the sample was 16.8 (\pm 2.5) years. Age of onset was unrelated to eating disorder outcome measured either categorically ($t_{161}, 1.45$; $P=.15$) or continuously ($\Delta R^2, 0.00018$; $n, 163$; $P=.86$). Similarly, the age of participants when they entered the study was not associated with eating disorder outcome measured either categorically ($t_{171}, 0.21$; $P=.84$) or continuously ($\Delta R^2, 0.017$; $n, 173$; $P=.08$).

Duration of symptoms at baseline was measured as the difference between age of onset and age of presentation. Correcting for outliers, mean (\pm SD) duration of symptoms prior to presentation was 5.9 (\pm 3.6) years. Duration of eating disorder symptoms at baseline was significantly correlated with eating disorder outcome (categorical eating disorder outcome: $t_{163}, 3.05$; $P=.003$; $P<.01$; continuous eating disorder outcome: $\Delta R^2, 0.061$; $n, 165$; $P=.001$; $P<.01$). Longer duration of symptoms at baseline predicted worse eating disorder outcome.

Baseline severity of eating disorder symptoms was measured by a section of the Eating Disorders Questionnaire that assessed the frequency of disordered eating behaviors within the month prior to assessment. Baseline

Table 2. Matched Pair *t*-Test Comparisons of Weight Variables Measured at Baseline and Follow-up

Characteristics	Baseline		Follow-up		<i>t</i>	<i>df</i>	<i>P</i>
	Mean	SD	Mean	SD			
Body mass index, kg/m ² *	21.20	2.72	22.09	3.61	3.80	148	<.001†
Weight, kg	58.28	8.52	60.73	10.88	3.48	149	.001‡
Height, cm	165.94	6.12	165.86	6.12	0.37	158	.71
Desired weight, kg§	53.09	5.23	56.51	6.18	9.56	140	<.001†
Highest weight, kg	66.38	11.43	69.79	13.18	4.26	152	<.001†
Lowest weight, kg	50.91	7.38	50.91	8.07	0.00	148	1.00

*At baseline assessment, all height and weight data were collected as actual measurements. At follow-up assessment, 51% of weight and 50% of height data were collected as actual measurements and remaining data were gathered through self-report on the Eating Disorders Questionnaire. Among subjects with both actual and self-reported measures of weight, correlation $r, 0.99$; $N, 60$; and $P<.001$; matched-pair $t, 1.68$; $df, 59$; and $P, .10$.

†Significance level after Dunn test correction for family-wise error rate, $P<.001$.

‡Significance level after Dunn test correction for family-wise error rate, $P<.01$.

§Desired weight represents the weight an individual wished to be at the time of assessment.

||Highest and lowest weights represent the highest and lowest weights the individual achieved between the age of 18 and the assessment indicated.

severity showed no significant relationship to eating disorder outcome measured either categorically ($t_{159}, 0.55$; $P=.58$) or continuously ($\Delta R^2, 0.0035$; $n, 161$; $P=.44$).

Comorbid Psychopathology

Participants meeting lifetime criteria for anorexia nervosa from SCID-I/P interviews and reporting onset of anorexia nervosa prior to onset of bulimia nervosa were compared with subjects without a history of anorexia nervosa. Consistent with findings from other studies, history of anorexia nervosa did not predict eating disorder outcome measured either categorically or continuously ($\chi^2_1, 1.79$; $P=.18$; $F_{1,119}, 1.45$; $P=.23$).

Other lifetime Axis I disorders determined from SCID-I/P interviews were collapsed into 4 categories (ie, mood disorders, anxiety disorders, substance use disorders, and impulse control disorders). The presence of Axis I disorders was not systematically assessed at baseline, so conclusions regarding the predictive association between lifetime Axis I diagnoses and eating disorder outcome must be tempered because these disorders may not have been diagnosable at presentation. **Table 3** presents associations between lifetime Axis I disorders and eating disorder outcome. Lifetime substance use disorders was the only Axis I category significantly associated with the number of months since last disordered eating behavior.

Associations between eating disorder outcome and baseline assessments of depression, anxiety, substance use problems, and impulse control problems are also presented in Table 3. Both follow-up assessments of lifetime Axis I disorders and baseline assessments of comorbid psychopathology suggest that substance use disorders predicted more recent binge eating and/or purging episodes at follow-up.

Baseline measures of impulse control problems did not predict eating disorder outcome measured either categori-

Table 3. Associations Between Lifetime Axis I Disorders and Eating Disorder Outcome

Characteristics	Eating Disorder Outcome					
	Categorical				Continuous	
	Remission, No. (%)	Disordered Eating, No. (%)	χ^2_1	P	F (df)	P
Structured Clinical Interview for <i>DSM-IV</i> lifetime diagnoses						
Mood	76 (62.8)	37 (71.2)	1.51	.22	4.91 (1, 169)	.028
Anxiety	36 (29.8)	18 (34.6)	0.40	.53	0.05 (1, 170)	.83
Substance use	64 (53.8)	37 (74.0)	5.99	.01†	11.14 (1, 166)	.001‡
Impulse control	20 (16.5)	11 (21.2)	0.53	.47	1.81 (1, 170)	.18
Psychopathology at baseline						
Substance use§	10 (19.2)	7 (43.8)	3.92	.05	14.02 (1, 65)	.001
Impulse control	31 (46.3)	18 (58.1)	1.18	.28	0.94 (1, 95)	.33
	Mean (SD)	Mean (SD)	t (df)	P	ΔR^2 (N)	P
Psychopathology at baseline						
Depression¶	13.2 (7.7)	13.3 (8.0)	0.06 (171)	.95	0.011 (173)	.17
Anxiety#	7.7 (4.6)	7.6 (6.1)	0.10 (101)	.92	0.015 (103)	.22

*The number of patients in remission was 121; those who still exhibited disordered eating, 52.

†Significance level after Dunn test correction for family-wise error rate, $P < .05$.

‡Significance level after Dunn test correction for family-wise error rate $P < .01$.

§Assessed from an earlier version of the Eating Disorders Questionnaire completed by the 2- to 5-year follow-up subsample.

||Assessed from the Eating Disorders Questionnaire in the total sample by distinguishing between women reporting at least 2 vs none of the following behaviors: suicide attempts, severe autoaggression, stealing, and alcohol abuse.

¶Assessed from the Hamilton Depression Rating Scale for the total sample.

#Assessed from the Hamilton Anxiety Rating Scale for the treatment outcome subsample.

Table 4. Treatment and Eating Disorder Outcome

Characteristics	Eating Disorder Outcome					
	Categorical*				Continuous	
	Remission, No. (%)	Disordered Eating, No. (%)	χ^2_1	P	t_{171}	P
Treatment received						
Medication	84 (69.4)	43 (82.7)	3.28	.07	2.72	.007†
Therapy	115 (95.0)	49 (94.2)	0.05	.83	0.63	.53
Current treatment						
Medication	26 (21.4)	23 (44.2)	9.27	.002‡	3.07	.002‡
Therapy	14 (11.5)	16 (30.8)	9.35	.002‡	3.20	.002‡

*The number of patients in remission was 121; those who still exhibited disordered eating, 52.

†Significance level after Dunn test correction for family-wise error rate, $P < .05$.

‡Significance level after Dunn test correction for family-wise error rate, $P < .01$.

cally or continuously. Unfortunately, no other measures of personality variables were available from baseline assessments. The association between Multidimensional Personality Questionnaire-Control Scale scores and eating disorder outcome suggested that personality may be a significant predictor of outcome (categorical: $t_{153}, 3.27$; $P = .001$; $P < .01$ and continuous: $r, 0.22$; $n, 155$; $P = .006$; $P < 0.05$); however, these analyses were cross-sectional and no conclusions can be drawn regarding the extent that personality variables predict eating disorder outcome.

Treatment

Participants from the treatment outcome study⁷ were originally stratified according to Hamilton Depression Rating Scale scores and then randomly assigned to 1 of 4 treat-

ment cells: imipramine hydrochloride, placebo, imipramine hydrochloride combined with intensive group cognitive behavioral therapy, or placebo combined with intensive group cognitive behavioral therapy. Participants from the 2- to 5-year follow-up study⁶ participated in an intensive outpatient group therapy similar but not identical to that employed in cells 3 and 4 in the treatment outcome study.⁷ For the total sample, baseline treatment conditions were defined by the 4 cells for women in the treatment outcome study⁷ and a fifth cell (group treatment) for the 2- to 5-year follow-up sample.⁶ No significant differences among treatment conditions were found for eating disorder outcome measured either categorically ($\chi^2_4, 4.26$; $P = .37$) or continuously ($F_{4,167}, 0.18$; $P = .95$).

Because subjects were not in a closed follow-up period, whether subjects had ever received psychotropic medi-

cation or psychotherapy was assessed during SCID-I/P interviews. **Table 4** presents findings for associations between treatment history and eating disorder outcome. The percentages of women receiving psychotherapy and medication at some time in their history were high and reduced the likelihood of finding associations with eating disorder outcome; however, women who received medication were more likely to report recent disordered eating behaviors at a trend level. Subsequent prescription of medication may have been driven by continued symptoms rather than medication producing worse eating disorder outcome. Supporting this interpretation, women with worse eating disorder outcome were more likely to be in therapy or receiving psychoactive medication at the time of follow-up assessment (Table 4).

COMMENT

More than a decade following presentation, 11% of this sample met full criteria for bulimia nervosa and 0.6% met full criteria for anorexia nervosa. An additional 18.5% met criteria for eating disorder not otherwise specified, and 69.9% were in remission. The findings suggested that the number of women who met full criteria for bulimia nervosa declined as follow-up duration increased, but approximately 30% continued to engage in recurrent binge eating or purging behaviors (incidence rate, 0.026 cases per person-years). A history of substance use disorders and longer duration of the disorder at presentation predicted worse outcome. Baseline treatment condition was unrelated to eating disorder outcome.

Strengths of our study include its duration and sample size, as it offers, to our knowledge, the longest follow-up period in the English-language literature on bulimia nervosa with the second largest sample size. The ascertainment (91%) and participation (80%) rates exceed the average for these parameters among existing bulimia nervosa follow-up studies. The assessment employed psychometrically validated instruments that largely displayed high α and κ reliabilities. Analyses were conducted with conservative *P* value corrections, and many assumptions underlying various statistical analyses were investigated for their validity.

There are many weaknesses in the current study. Approximately 20% of the sample did not participate in follow-up assessment, and half of these women could not be located. To reduce the participation requirement (and increase participation rates), more extensive personality assessments were excluded from the protocol. Given the failure to determine the effect of impulse control problems on eating disorder outcome, more comprehensive personality assessment would be an important area for future investigations. In addition, our sample was not ethnically diverse and did not include male subjects, so our findings may not apply to individuals who differ demographically from the present sample. Similarly, our results are generalizable only to clinical samples because participants were initially seeking treatment.

The purpose of this study was to determine the long-term outcome associated with bulimia nervosa and predictors of long-term outcome. Mortality was a rare outcome associated with bulimia nervosa occurring in only

1 (0.5%) of 200 subjects ascertained for this study. This finding is surprising given the 5.9% crude mortality rate associated with anorexia nervosa¹⁵ and the percentage of this sample with lifetime affective disorder diagnoses. The portion of this sample meeting criteria for a primary mood disorder demonstrated an unusually low rate of mortality due to suicide (0.9%) when compared with rates in follow-up studies of patients with mood disorders^{16,17}

Depending on definitions, 42% to 47% of women were in full remission. This range approaches the 50% value predicted by our review.³ The percentage of women still meeting full criteria for bulimia nervosa (11%) was close to the 9% reported in the 10-year follow-up study by Collings and King⁵ but lower than predicted. Considering results across studies, the percentage of women in full remission appears to reach an asymptote near 50%, but the percentage of women continuing to meet full criteria for bulimia nervosa seems to decline as the length of follow-up increases. Supporting this interpretation, the duration of follow-up was significantly associated with recency of disordered eating behaviors.

A lifetime history of substance use disorders and baseline reports of alcohol and other drug problems were predictive of long-term outcome measured continuously. This replicated findings of 1 other study¹⁸ but contradicted findings of several other studies.¹⁹⁻²² The mean sample size of studies showing no association ($n=50$) was less than the mean of studies showing a significant association ($n=185$). Possibly, studies failing to find an association may have lacked sufficient statistical power to reveal the predictive significance of substance use disorders.

Longer duration of bulimia nervosa before presentation predicted worse eating disorder outcome similar to findings of previous studies^{23,24}; however, other studies have found no predictive significance for this variable.^{5,19,22} Further replication of this finding could support the existence of a subgroup of individuals for whom bulimia nervosa runs a chronic course. Alternatively, duration may represent time between onset of symptoms and treatment seeking; seeking treatment soon after the onset of symptoms may improve prognosis. Unfortunately, we did not know the age when treatment was first sought and cannot directly test this hypothesis.

Lifetime history of major depression and the degree of depressive symptoms at baseline were not predictive of eating disorder outcome. This finding corroborated results of several other studies^{6,20,25-27}; however, baseline diagnoses of major depression were not made, and we cannot make conclusions concerning the effect of major depression at presentation. The failure of baseline impulse control problems to predict outcome contradicted results from previous investigations.^{18,21,28,29} If scores from the Multidimensional Personality Questionnaire-Control Scale represent a trait rather than a state, then impulsivity does influence eating disorder outcome; however, impulsivity as a trait was not assessed at baseline. Therefore, the effect of personality characteristics on bulimia nervosa outcome remains unclear for this sample.

Baseline treatment interventions were not associated with the long-term outcome despite significant dif-

ferences in treatment response favoring cognitive behavioral therapy.⁷ Although subjects were not in a closed follow-up period, it seems unlikely that individuals in the placebo- or medication-only groups later received manual-based cognitive behavioral therapy. A recent study³⁰ reported that among women with bulimia nervosa screened for treatment study participation, only 6.9% had ever received cognitive behavioral therapy. A lack of detailed longitudinal data on treatment for this sample makes it difficult to draw firm conclusions concerning the association between treatment and long-term outcome.

Although baseline treatment conditions were not directly associated with long-term outcome in this study, this does not rule out an effect for treatment interventions. Fairburn et al⁴ found significant differences in outcome among women based on their treatment condition an average of 5.8 years after presentation. Treatments may speed eventual recovery from bulimia nervosa, but their effects may be masked by the natural course of the disorder (ie, women who recover without the benefit of the most efficacious treatment). Such effects would be undetectable by the snapshot of outcome presented by the current study.

This study demonstrates that women improve over time as fewer continue to meet full criteria for bulimia nervosa. A large proportion of women, however, suffer from threshold and subthreshold eating disorders more than a decade after diagnosis. Future research in this area would benefit from exploring the association between treatment and course of bulimia nervosa and by continuing to develop increasingly efficacious treatment interventions.

Accepted for publication October 15, 1998.

This study was supported in part by a McKnight Center grant for Eating Disorders Research, Minneapolis, Minn (Dr Mitchell); Obesity Center grant P30 DK50456 from the National Institutes of Health, Bethesda, Md (Dr Mitchell); a research training grant from the National Institute of Mental Health, Bethesda, Md (Dr Iacono); dissertation grants from the American Psychological Association, Washington, DC and Minnesota Women Psychologists' Association, Minneapolis, Minn (Dr Keel); and a dissertation fellowship from the University of Minnesota, Minneapolis (Dr Keel).

Presented in part at the third meeting of the Eating Disorder Research Society, Albuquerque, NM, November 14-16, 1997.

We thank Kelly Ball, Matthew Biaocchi, Ross D. Crosby, PhD, Elke D. Eckert, MD, Sara Engbloom, Debbie Glotter, Gretchen Goff, Jane Harper, Dorothy Hatsukami, PhD, Gloria R. Leon, PhD, Carol B. Peterson, PhD, Claire Pomeroy, Richard L. Pyle, MD, Joshua S. Rodefer, PhD, and Robert Zimmerman for their help with this study.

Reprints: Pamela K. Keel, PhD, 1102 William James Hall, Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138.

REFERENCES

- Russell G. Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychol Med*. 1979;9:429-448.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
- Keel PK, Mitchell JE. Outcome in bulimia nervosa. *Am J Psychiatry*. 1997;154:313-321.
- Fairburn CG, Norman PA, Welch SL, O'Connor ME, Doll HA, Peveler RC. A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments. *Arch Gen Psychiatry*. 1995;52:304-312.
- Collings S, King M. Ten-year follow-up of 50 patients with bulimia nervosa. *Br J Psychiatry*. 1994;164:80-87.
- Mitchell JE, Pyle RL, Hatsukami D, Goff G, Glotter D, Harper J. A 2-5 year follow-up of patients treated for bulimia. *Int J Eat Disord*. 1988;8:157-165.
- Mitchell YE, Pyle RL, Eckert ED, Hatsukami D, Pomeroy C, Zimmerman R. A comparison study of antidepressants and structured intensive group psychotherapy in the treatment of bulimia nervosa. *Arch Gen Psychiatry*. 1990;47:149-157.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association; 1980.
- Mitchell YE, Hatsukami D, Eckert E, Pyle R. Eating Disorders Questionnaire. *Psychopharmacol Bull*. 1985;21:1025-1043.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56-62.
- Hamilton M. The assessment of anxiety by rating. *Br J Med Psychol*. 1959;32:50-55.
- Tellegen A. *Brief Manual for the Differential Personality Questionnaire*. Minneapolis: University of Minnesota; 1982.
- Cooper PJ, Taylor MJ, Cooper Z, Fairburn CG. The development and validation of the Body Shape Questionnaire. *Int J Eat Disord*. 1987;6:485-494.
- First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition (SCID-I/P, Version 2.0)*. New York, NY: New York State Psychiatric Institute Biometrics Research Dept; 1995.
- Sullivan PF. Mortality in anorexia nervosa. *Am J Psychiatry*. 1995;152:1073-1074.
- Sainsbury P. Suicide: epidemiology and relationship to depression. In: Wing JKW, ed. *Cambridge Handbook of Psychiatry: Psychoses of Uncertain Etiology*. Cambridge, England: Cambridge University Press; 1982:134-140.
- Harris E, Barraclough B. Suicide as an outcome for mental disorders. *Br J Psychiatry*. 1997; 170:205-228.
- Fichter MM, Quadflieg N, Rief W. Longer-term course (6-year) course of bulimia nervosa. *Neuropsychopharmacology*. 1994; 10:772S.
- Abraham SF, Mira M, Llewellyn-Jones D. Bulimia: a study of outcome. *Int J Eat Disord*. 1983;2:175-180.
- Fallon BA, Walsh T, Sadik C, Saoud JB, Lukaski V. Outcome and clinical course in inpatient bulimic women: a 2- to 9-year follow-up study. *J Clin Psychiatry*. 1991;52:272-278.
- Herzog T, Hartmann A, Sandholz A, Stammer H. Prognostic factors in outpatient psychotherapy of bulimia. *Psychother Psychosom*. 1991;56:48-55.
- Mitchell JE, Davis L, Goff G, Pyle R. A follow-up study of patients with bulimia. *Int J Eat Disord*. 1986;5:441-450.
- Fahy TA, Russell GFM. Outcome and prognostic variables in bulimia nervosa. *Int J Eat Disord*. 1993;14:135-145.
- Hsu LKG, Holder D. Bulimia nervosa: treatment and short-term outcome. *Psychol Med*. 1986;16:65-70.
- Herzog DB, Keller MB, Lavori PW, Ott IL. Short-term prospective study of recovery in bulimia nervosa. *Psychiatry Res*. 1988;23:45-55.
- Johnson-Sabine E, Reiss D, Dayson D. Bulimia nervosa: a 5-year follow-up study. *Psychol Med*. 1992;22:951-959.
- Walsh BT, Hadigan CM, Devlin MJ, Gladis M, Roose SP. Long-term outcome for antidepressant treatment for bulimia nervosa. *Am J Psychiatry*. 1991; 148:1206-1212.
- Fichter MM, Quadflieg N, Rief W. Course of multi-impulsive bulimia. *Psychol Med*. 1994;24:591-604.
- Johnson C, Tobin DL, Dennis A. Differences in treatment outcome between borderline and nonborderline bulimics at one-year follow-up. *Int J Eat Disord*. 1990; 9:617-627.
- Crow SJ, Mitchell JE, Mussell MP, Peterson CB, Knopke A. Adequacy of prior treatment for bulimia nervosa. Paper presented at: Eating Disorders Research Society; November 15, 1996; Pittsburgh, Pa.