

Incidence and outcomes of bulimia nervosa: a nationwide population-based study

A. Keski-Rahkonen^{1,2,3*}, H. W. Hoek^{3,4,5}, M. S. Linna¹, A. Raevuori^{1,3}, E. Sihvola^{1,2}, C. M. Bulik⁶,
A. Rissanen² and J. Kaprio^{1,7,8}

¹ Department of Public Health, University of Helsinki, Finland

² Department of Psychiatry, Helsinki University Central Hospital, Finland

³ Department of Epidemiology, Columbia University, USA

⁴ Parnassia, The Hague, The Netherlands

⁵ Department of Psychiatry, Groningen University, The Netherlands

⁶ Departments of Psychiatry and Nutrition, University of North Carolina at Chapel Hill, USA

⁷ Department of Mental Health and Alcohol Research, National Public Health Institute, Helsinki, Finland

⁸ Academy of Finland, Center of Excellence in Complex Disease Genetics, Finland

Background. Little is known about the epidemiology of bulimia nervosa outside clinical settings. We report the incidence, prevalence and outcomes of bulimia nervosa using for the first time a nationwide study design.

Method. To assess the incidence and natural course and outcomes of DSM-IV bulimia nervosa among women from the general population, women ($n=2881$) from the 1975–79 birth cohorts of Finnish twins were screened for lifetime eating disorders using a two-stage procedure consisting of a questionnaire screen and the Structured Clinical Interview for DSM-IV (SCID). Clinical recovery was defined as 1-year abstinence from bingeing and purging combined with a body mass index (BMI) ≥ 19 kg/m².

Results. The lifetime prevalence of DSM-IV bulimia nervosa was 2.3%; 76% of the women suffered from its purging subtype and 24% from the non-purging subtype. The incidence rate of bulimia nervosa was 300/100000 person-years at the peak age of incidence, 16–20 years, and 150/100000 at 10–24 years. The 5-year clinical recovery rate was 55.0%. Less than a third of the cases had been detected by health-care professionals; detection did not influence outcome. After clinical recovery from bulimia nervosa, the mean levels of residual psychological symptoms gradually decreased over time but many women continued to experience significantly more body image problems and psychosomatic symptoms than never-ill women.

Conclusions. Few women with bulimia nervosa are recognized in health-care settings. Symptoms of bulimia are relatively long-standing, and recovery is gradual. Many clinically recovered women experience residual psychological symptoms after attaining abstinence from bingeing and purging.

Received 9 November 2007; Revised 19 May 2008; Accepted 24 May 2008; First published online 8 September 2008

Key words: Bulimia nervosa, epidemiology, incidence, natural course, outcome, prevalence.

Introduction

Bulimia nervosa affects adolescent and young adult women (Walsh & Devlin, 1998; Becker *et al.* 1999; Fairburn & Harrison, 2003). Few studies exist on its incidence (Hoek *et al.* 1995; Currin *et al.* 2005; Van Son *et al.* 2006) and course (Fairburn *et al.* 2000; Lewinsohn *et al.* 2000; Wade *et al.* 2006).

Keel & Mitchell (1997) summarized 88 outcome studies of bulimia nervosa. Five to 10 years after presentation, approximately 50% of women initially

diagnosed with bulimia nervosa had fully recovered from their disorder. A chronic form of bulimia nervosa was relatively common; about 20% continued to meet full criteria for bulimia nervosa at the end of the follow-up period. However, comparisons across different outcome studies were difficult because of varying diagnostic categories over time, lack of consensus about definitions of recovery, high drop-out rates, and varying lengths of follow-up time. Long-term follow-up studies of bulimia nervosa also primarily reflected clinical case series or catchment area studies (Keel & Mitchell, 1997; Fairburn *et al.* 2000; Berkman *et al.* 2007). Only a minority of women with bulimia nervosa are detected by primary care or specialized mental health care (Hoek & van Hoeken, 2003).

* Address for correspondence: Dr A. Keski-Rahkonen, Embaixada de Finlândia, SES Av. das Nações, Qd 807, Lt 7, CEP 70417-900, Brasília DF, Brasil.

(Email: anna.keski-rahkonen@helsinki.fi)

The aim of this study was to overcome some of the limitations of existing studies by addressing the incidence, prevalence and outcomes of bulimia nervosa using for the first time a naturalistic setting, a nationwide non-clinical study of the 1975–79 twin birth-cohorts in Finland. In addition, we examined the behavioral, psychological and social outcomes of bulimia nervosa, comparing women with bulimia nervosa to their unaffected co-twins and unrelated healthy control women.

Method

FinnTwin16 birth cohorts

This nationwide longitudinal cohort study of health behaviors in twins and their families (Kaprio *et al.* 2002) identified almost all of the live twin births in 1975–79 from the central population register of Finland. The twins and their parents were sent baseline self-report questionnaires when the twins were 16 years old. Subsequently, three waves of follow-up questionnaires were mailed to the twins when they were aged 17, 18 and 21–27. Data collection and analysis was approved by the ethics committee of the Department of Public Health, University of Helsinki and by the internal review boards of Indiana University and Columbia University.

Questionnaire screening for eating disorders

At age 21–27 years (mean age 24.4 years, s.d. = 0.9), the wave four questionnaire ($n=2881$) yielded a sample of 2545 responders, of whom 868 were monozygotic, 765 same-sex dizygotic, 793 females from opposite-sex dizygotic pairs, and 119 of unknown zygosity. The screening included self-report questions addressing current height and weight, ideal weight, and minimum and maximum weight at current height. Purging (vomiting or laxative, diuretic, or enema use) was assessed by a multiple-choice question that differentiated current repeated purging (for at least 3 months) from past repeated purging, an occasional try, and never purging. Other bulimic behaviors and body image were assessed using three subscales of the Eating Disorder Inventory (EDI: Body Dissatisfaction, Drive for Thinness, and Bulimia) (Garner, 1991). We have reported further details of the screening elsewhere (Keski-Rahkonen *et al.* 2006). All screen-positive women ($n=292$), their screen-negative female co-twins ($n=130$) and 210 randomly selected screen-negative women were invited to participate in the second stage of screening, diagnostic telephone interviews (Keski-Rahkonen *et al.* 2006, 2007).

Diagnostic interviews

Five experienced clinicians (four MDs and one RN) from the Eating Disorder Unit of Helsinki University Central Hospital received detailed training in conducting the short version of the Structured Clinical Interview for DSM-IV (SCID; First *et al.* 2003). The interviews were administered by telephone to obtain current and lifetime diagnoses of anorexia nervosa, bulimia nervosa and binge-eating disorder, to define age of the participants at the first and last manifestations of these illnesses, and to establish the temporal sequence of these diagnoses and their respective time courses. Inter-rater agreement for diagnosis was good (mean $\kappa=0.87$, range 0.64–1.00). A.K.-R. and E.S. supervised interviewers throughout the study and helped to solve diagnostic problems, such as concerns about diagnostic threshold and differential diagnosis.

The overall interview participation rate was 85.2% (90.1% for the screen-positive, 76.2% for the screen-negative female co-twins, and 84.8% for the random screen-negative sample). Only 12 individuals (1.9%) refused to be interviewed; the rest could not be reached because of difficulties in obtaining their current telephone number.

Definition of disorder

We used two definitions of bulimia nervosa. Few meaningful differences exist between women who engage in bulimic behaviors twice a week and those who engage in such behaviors less frequently (Garfinkel *et al.* 1995; Sullivan *et al.* 1998). Thus, the DSM-IV frequency criterion C that requires a symptom frequency of at least twice a week is currently controversial, and its future in DSM-V is questionable (Wilfley *et al.* 2007). In this study, the current DSM-IV definition (the binge eating and compensatory behaviors both occur at least twice a week for 3 months) of bulimia nervosa is termed from here on 'BN narrow'. This definition yielded 42 cases. The broader definition, where symptom frequency was relaxed to once a week, 'BN broad', yielded 59 cases.

From previous literature, it is obvious that only a small proportion of cases of bulimia nervosa are detected in health-care settings (Hoek & van Hoeken, 2003; Hudson *et al.* 2007). Thus, we assessed whether participants of this study had received an eating disorder diagnosis from a health professional prior to this study.

Five-year clinical recovery rates and outcomes after clinical recovery

For each case of bulimia nervosa, the interviewers determined the last age of occurrence for any eating

disorder symptom, including the psychological features of self-evaluation of body image, self-esteem and anxiety. Definitions of remission and recovery remain conflicting for bulimia nervosa; although many clinical trials use far less stringent criteria of remission and recovery, the likelihood of symptom recurrence only decreases sharply after 1 year of abstinence from bingeing and purging (Field *et al.* 1997). We therefore defined 'clinical recovery' as the abstinence from the core behavioral eating disorder symptoms, bingeing and purging, for at least 1 year prior to assessment. We also required that women in clinical recovery had a self-reported body mass index (BMI) ≥ 19 kg/m² to ensure that individuals with DSM-IV or subthreshold anorexia nervosa were not classified as recovered. Agreement between self-reported and measured BMI in the women from this birth cohort was excellent, $r=0.92$ (Schousboe *et al.* 2003).

The 5-year clinical recovery rate was defined as the proportion who reached clinical recovery within 5 years after disorder onset. Outcomes after clinical recovery were addressed separately using self-report measures in two domains. First, psychological aspects of eating disorders were measured by four scales of the EDI: Drive for Thinness, Bulimia, Body Dissatisfaction, and Perfectionism (Garner, 1991), by Spielberger's State-Trait Anxiety Inventory (Spielberger *et al.* 1983), and by Rosenberg's Self-Esteem Scale (Rosenberg, 1989). Second, social functioning was measured by occupation, education, marriage or cohabitation, and parity.

Using these psychological and social outcome measures, we also compared women with bulimia nervosa to their unaffected female co-twins and a random sample of unrelated unaffected women ($n=134$) (referred to from here on as 'healthy women'). We defined 'unaffected' as the absence of eating disorder symptoms in both the questionnaire screening and the SCID. In the questionnaire, unaffected individuals did not report anorexia nervosa, bulimia nervosa, purging behaviors or current low BMI (<17.6 kg/m²), did not score above the 97th percentile in EDI – Bulimia, and did not have a combination of low BMI (<17.6 kg/m²) with EDI – Body Dissatisfaction score >70 th percentile or EDI – Drive for Thinness score ≥ 90 th percentile (Keski-Rahkonen *et al.* 2006). In addition, the SCID could not have yielded positive diagnoses for DSM-IV bulimia nervosa, anorexia nervosa, binge-eating disorder, or eating disorder not otherwise specified. After the exclusion of male or non-responding co-twins and the exclusion of one female co-twin with DSM-IV anorexia nervosa and three co-twins with subthreshold bulimic symptoms, 19 unaffected female co-twins remained.

Statistical analysis

Lifetime prevalence was calculated by dividing the number of lifetime prevalent cases of bulimia nervosa by the total number of women who responded in the eating disorder screen at age 21–27 years ($n=2545$).

Incidence rates were calculated for the peak age at risk in this sample (16–20 years), and also for the age group 15–19 years to facilitate comparisons with previous studies. We also calculated incidence rates for the full age range at risk (10–24 years). The numerator was the number of incident cases of bulimia nervosa detected in the given age range and the denominator was the number of person-years at risk in the same age range. As not all of the women had reached age 24 (the mean age at interview was 25.9 years, s.d. = 1.5 years, range 22–29 years), for the calculation of the 10–24 years incidence, we deducted 1160 person-years from the denominator to account for those who were <25 years at the time of the interview.

The 5-year clinical recovery rate was computed using the Kaplan–Meier survival method. The log-rank test was used to compare survival curves of different diagnostic subgroups. We used Cox proportional hazard models to test the effect of age of onset, eating disorder subtype and co-morbidity on recovery rates.

Psychological outcomes after clinical recovery were analyzed by calculating mean psychological outcome scores and 95% confidence intervals (CIs) of means for cases of bulimia nervosa who were currently ill, cases in clinical recovery for >5 years, and the two control groups (unaffected co-twins and healthy women). The statistical power was very limited for between-group comparisons; therefore, these analyses were mainly exploratory, and the absence of statistical significance (i.e. overlapping CIs) cannot be interpreted as lack of association.

For analyses of social outcomes, we used analysis of variance for continuous and logistic regression for categorical outcome measures. To account for clustered sampling within the twin pair, p values and CIs were adjusted using standard procedures for survey data (Williams, 2000). All analyses were conducted using Stata Statistical Software release 9.2 (Stata Corporation, College Station, TX, USA).

Results

Prevalence and incidence

The lifetime prevalence of BN narrow was 1.7% and BN broad 2.3% (Table 1). Only about a third of the lifetime prevalent cases of bulimia nervosa had been detected by the health system (Table 1). The peak age group at risk for bulimia nervosa comprised women

Table 1. The prevalence, incidence and detection rate of narrow and broad bulimia nervosa (with 95% confidence intervals)

	Narrow bulimia nervosa ^a	Broad bulimia nervosa ^b
Number of affected individuals in cohort	42	59
Number of affected twin pairs in cohort	36	51
Discordant	30	43
Concordant	6	8
Lifetime prevalence (%)	1.7 (1.2–2.1)	2.3 (1.7–2.9)
Incidence (10–24 years), per 100000 person-years	100 (75–140)	150 (120–200)
Incidence (16–20 years), per 100000 person-years	200 (130–290)	300 (210–400)
Incidence (15–19 years), per 100000 person-years	210 (130–290)	280 (190–370)
Disorder detected by health-care professionals prior to this study (%)	38.1	32.2

^aSymptom frequency at least twice a week for 3 months.

^bSymptom frequency at least once a week for 3 months.

aged 16–20 years (Fig. 1); their incidence rate was twice the incidence rate of the age group 10–24 years (Table 1).

Subtypes of bulimia nervosa

We found that 79% of women with BN narrow and 77% of women with BN broad engaged in purging (i.e. self-induced vomiting, laxative, diuretic, or diet pill use). The rest suffered from the non-purging subtype of the illness (i.e. resorted to excessive exercise, fasting, or repeated dieting to compensate for binge eating).

History of anorexia nervosa

Crossover from DSM-IV anorexia nervosa to bulimia nervosa took place in 31% of the BN narrow and 24% of the BN broad probands. No one crossed over from bulimia nervosa to anorexia nervosa.

Five-year clinical recovery rates

The 5-year recovery rate from BN narrow was 57% and from BN broad, 55%; overall, 64% of women with BN narrow and 61% of women with BN broad reached clinical recovery during the study time. The recovery rates did not vary between women with detected and undetected bulimia nervosa ($p=0.49$ for BN narrow and $p=0.55$ for BN broad) (Fig. 2). Anorexia nervosa prior to onset of bulimia nervosa marginally significantly decreased the likelihood of recovery from BN narrow within 5 years ($p=0.07$), but the same trend was less apparent for BN broad ($p=0.18$). The subtype of bulimia nervosa (purging versus non-purging) did not influence outcome ($p=0.27$ for BN

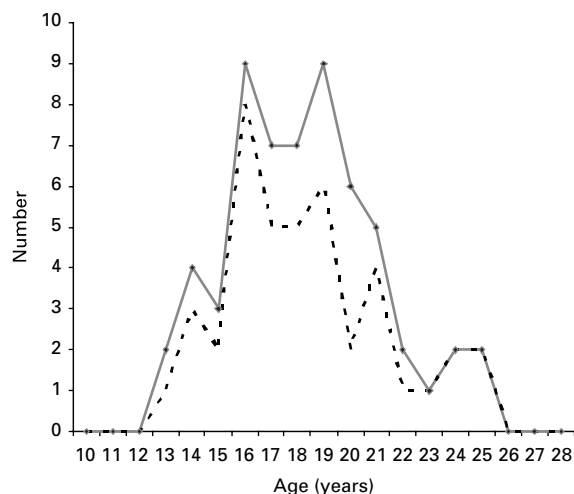


Fig. 1. Ages of onset of bulimia nervosa among women from the Finnish twin cohorts ($n=59$). BN narrow (---) signifies DSM-IV bulimia symptom frequency at least twice a week for 3 months; in BN broad (—), symptom frequency is relaxed to once a week.

narrow and $p=0.15$ for BN broad), nor did age of onset significantly affect the likelihood of recovery from bulimia nervosa ($p=0.32$ for both BN narrow and BN broad).

Outcomes of individuals with persisting symptoms

Of women with BN broad not classified as recovered, one had been in full remission for just under 1 year. A further 22 were still clearly symptomatic at the time of the interview, but only five (22.7%) still met the narrow DSM-IV criteria for bulimia nervosa and one

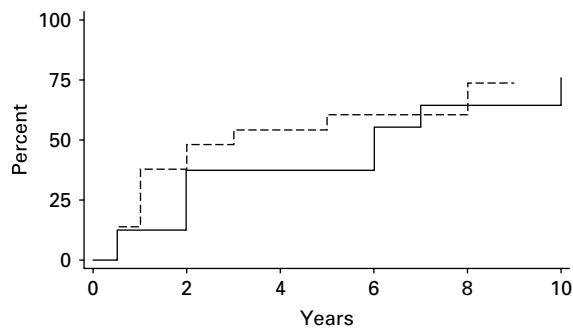


Fig. 2. Recovery rates from DSM-IV bulimia nervosa: women whose disorder was detected (—) versus undetected (---) by health professionals did not differ in their outcomes (log-rank test $p=0.49$).

(4.5%) met DSM-IV criteria for binge-eating disorder. Five women (22.7%) experienced residual symptoms of purging-type bulimia nervosa and two (9.1%) residual symptoms of non-purging-type bulimia nervosa; they had succeeded in reducing the frequency of binges or compensatory behaviors to less than weekly, and some had also reached periods of full or partial remission from some of the compensatory behaviors. Four women (18.2%) continued to have binges or regular overeating, but no more compensated for them; subsequently, they struggled with weight problems. Five women (22.5%) were unwilling to give specific information about their current symptoms, although it was apparent that they still suffered from some form of eating disorder not otherwise specified.

Outcomes after clinical recovery

The mean current BMI of BN narrow probands in clinical recovery was 22.7 (95% CI 21.3–24.0) kg/m^2 , and 22.9 (95% CI 21.8–23.9) kg/m^2 for BN broad probands. Neither was significantly different from that of their unaffected co-twins [23.1 (95% CI 21.7–24.5) kg/m^2].

In exploratory analyses of psychological functioning during illness and recovery (Table 2), the mean symptom scores of women with bulimia nervosa became more like those of their unaffected co-twins over time. Because of the small sample sizes, these analyses were only conducted using the BN broad group. In Table 2, we give CIs of means to facilitate between-group comparisons; however, because of our limited sample size and power, overlapping CIs cannot necessarily be interpreted as lack of association. Nevertheless, short-term clinical recovery (i.e. abstinence from bingeing and purging for 1–5 years) was not reflected in measures of psychological functioning; compared to currently ill women, the mean levels of anxiety and perfectionism remained high, self-esteem

low, and measures of body image relatively impaired among women who had been in clinical recovery for ≤ 5 years. Psychological improvement was more marked when currently ill women were compared to those who had been in clinical recovery for >5 years. However, many women who had been recovered from bulimia nervosa for >5 years also experienced significantly more body image problems and psychosomatic symptoms than did healthy women (Table 2).

In terms of social functioning, women currently ill with bulimia nervosa were significantly less likely to be married or cohabitating than their healthy co-twins ($p=0.018$), but did not differ from their healthy co-twins in educational attainment, unemployment rates, and parity. We observed no statistically significant differences in educational attainment, unemployment rates, marriage or cohabitation, and parity between probands who had reached clinical recovery from bulimia nervosa and their healthy co-twins.

Unaffected co-twins versus healthy unrelated women

Because unaffected co-twins may share with their affected co-twins genetic and environmental factors that predispose them to eating disorders, we compared unaffected co-twins to healthy women from the general population. Indeed, the unaffected co-twins were statistically significantly more likely to experience perfectionism than healthy women and they reported slightly but not statistically significantly more body image problems (Table 2). The unaffected co-twins were on average 0.9 years younger and marginally more likely to have a university level education than healthy unrelated women. However, unaffected co-twins and healthy women did not differ in unemployment rates, marriage and cohabitation, and parity from healthy women.

Discussion

To our knowledge, this is the first nationwide population-based outcome study of bulimia nervosa. Less than a third of the cases of bulimia nervosa were recognized in health-care settings, and detection did not influence outcomes. Symptoms of bulimia were relatively long-standing: the 5-year recovery rate was 57% from BN narrow and 55% from BN broad, and many clinically recovered women remained symptomatic thereafter. Thus our study independently confirms findings from earlier community-based outcome studies of bulimia (Fairburn *et al.* 2000; Wade *et al.* 2006) in a nationwide setting using a different strategy of case ascertainment. Bulimia is clearly a chronic illness; many individuals experience periods of relapses and recurrences, and although only few individuals

Table 2. Disorder characteristics and psychosocial outcome measures: group means of women currently ill with broad bulimia nervosa^a, and ≤ 5 years and > 5 years after clinical recovery^b

	Currently ill (<i>n</i> = 23)	Recovered ≤ 5 years (<i>n</i> = 19)	Recovered > 5 years (<i>n</i> = 16)	Unaffected co-twins (<i>n</i> = 19)	Healthy women (<i>n</i> = 134)
Age of eating disorder onset	17.9 (16.3–19.6)	18.2 (16.7–19.7)	16.3 (15.6–17.1)	N.A.	N.A.
Age at interview	25.4 (24.7–26.1)	25.2 (24.6–25.9)	25.9 (24.9–26.8)	25.5 (24.8–26.2)	26.6 (26.2–26.9)
Minimum body mass index	18.8 (17.8–19.7)	19.6 (17.8–21.4)	18.1 (16.7–19.6)	20.3 (19.5–21.2)	19.7 (19.2–20.1)
Current body mass index	23.0 (21.8–24.1)	24.0 (21.3–26.7)	22.4 (21.1–23.7)	23.1 (21.8–24.4)	22.3 (21.6–23.0)
Eating Disorder Inventory					
Drive for Thinness	31.2 (28.7–33.7)	27.4 (23.0–31.9)	23.5 (20.0–27.0)	19.9 (17.0–22.9)	17.3 (16.0–18.5)
Body Dissatisfaction	32.3 (29.0–35.6)	30.9 (25.1–36.7)	31.1 (25.3–36.9)	25.9 (21.9–29.9)	23.1 (21.3–25.0)
Bulimia	22.8 (20.6–25.0)	21.8 (17.9–25.7)	14.9 (12.6–17.3)	11.6 (9.5–13.7)	10.6 (10.1–11.1)
Perfectionism	20.0 (17.5–22.5)	23.8 (29.9–26.7)	18.3 (15.6–20.9)	18.4 (16.1–20.7)	15.0 (14.2–15.9)
State-Trait Anxiety Inventory					
Trait Anxiety	47.2 (43.8–50.5)	49.7 (43.5–55.9)	42.2 (37.5–46.9)	39.1 (35.7–42.5)	37.9 (36.7–39.2)
State Anxiety	42.7 (37.9–47.6)	49.4 (42.1–56.6)	36.5 (31.0–42.0)	35.6 (31.2–40.0)	33.2 (31.6–34.8)
Psychosomatic symptoms	7.7 (6.2–9.1)	5.8 (4.5–7.1)	6.7 (5.3–8.0)	6.2 (4.6–7.7)	4.7 (4.2–5.2)
Rosenberg Self-Esteem Scale	18.8 (16.2–21.5)	15.5 (11.8–19.2)	20.8 (18.0–23.5)	22.3 (20.2–24.4)	22.6 (21.7–23.6)

95% Confidence intervals are given in parentheses, but because of power limitations, overlapping confidence intervals cannot necessarily be interpreted as lack of true between-group differences.

^a Symptom frequency at least once a week for 3 months.

^b Defined as abstinence from bingeing and purging for at least 1 year and a body mass index ≥ 19 kg/m².

met full criteria for bulimia nervosa 5 years after the onset of the illness, subthreshold and atypical forms of the illness were very common. Although comparisons with other psychiatric illnesses, such as major depressive disorder (Lehmann, 1983), are not clear-cut, it appears that numbers of individuals with chronic forms of these two illnesses are relatively similar, but the general 5-year prognosis of bulimia appears more favorable.

In previous community-based studies, the generally accepted prevalence of bulimia nervosa among young females has been about 1% (range 0–4.5%) (Hoek & van Hoeken, 2003). A recent representative survey of the adult US household population aged 18 and older (Hudson *et al.* 2007) found a DSM-IV bulimia nervosa lifetime prevalence of 1.5% among adult females. Lifetime prevalences depend on the age structure of

the sample studied; our higher lifetime prevalence may reflect the fact that we targeted a younger cohort of women in their twenties. Our population coverage may also have been better because our study was population registry based and the response rates were high; a large majority of our study participants were women who never sought professional help from health-care providers for their eating disorder symptoms.

Our incidence rates were considerably higher than those obtained from existing studies limited to health-care settings, probably because not only were we able to examine cases detected by health-care professionals (the 'detected' incidence rate) but also we could better approximate the 'true' incidence rate of bulimia nervosa. However, direct comparisons are difficult because the definition disorder and the target ages

vary widely across studies. Our incidence rate among women aged 15–19 was 280/100000, whereas a large community-based study in the USA reported an incidence of bulimia nervosa in 1980–90 of 125/100000 (Soundy *et al.* 1995). A Swedish study from the 1980s (Cullberg & Engström-Lindberg, 1988) found that the incidence of DSM-III bulimia was 65/100000 among women aged 16–24, which very closely resembles the ‘detected’ incidence rate in our study population. In primary care settings in the UK, the incidence of bulimia among the age group 10–19 years was 35.8/100000, with a clear decline in the incidence rates after 1996 (Currin *et al.* 2005). Our study was conducted during the same time period, and both our ‘detected’ and ‘true’ incidence rates were considerably higher. However, because there was little age heterogeneity in our sample, we could not examine time trends. Nevertheless, in Finland, incidence rates of bulimia are still high and the majority of cases receive neither diagnostic attention nor appropriate treatment from health-care professionals.

We found that the incidence of bulimia nervosa peaked at age 16–20 years, earlier than in many other existing epidemiological studies, probably because in our study the age of onset was not confounded by treatment delay. In eating disorders, this delay may be considerable even in countries with highly developed health systems, often 2–5 years (Willi *et al.* 1990; Milos *et al.* 2004), even when the illness is very severe. Common misconceptions about bulimia nervosa increase the stigma associated with the illness: women suffering from bulimia nervosa often hide their symptoms for long periods of time, and many also think that there may be no help available for their problems (Mond *et al.* 2006; Hepworth & Paxton, 2007).

However, our exploratory analyses revealed that after reaching clinical recovery, many women experienced multiple psychological symptoms, including body image problems, anxiety and low self-esteem. This may reflect the fact that recovery from bulimia nervosa is a very gradual process, and the period of vulnerability that follows abstinence from bingeing and purging may be longer than previously thought. It is also possible that bingeing and purging associated with bulimia nervosa may become egosyntonic over time and serve an anxiolytic function, and abstinence from bingeing and purging may result in a temporary resurgence of psychological suffering. In fact, Kaye *et al.* (1998) found that women recently recovered from bulimia nervosa suffered from more negative moods and obsessions with perfectionism and exactness compared to healthy women, and that these changes were paralleled by increased levels of the serotonin metabolite 5-hydroxyindoleacetic acid in cerebrospinal fluid. It is unclear whether these changes are permanent

(trait-related) or transient (state-dependent); in our sample, psychological improvement was more marked among women who had been in clinical recovery for longer periods of time, at least 5 years. It is also possible that our measure of choice, the EDI – Perfectionism subscale, is more sensitive to illness status (i.e. state-dependent) than other measures of perfectionism (Sutandar-Pinnock *et al.* 2003). Nevertheless, even among women in recovery for at least 5 years, body image problems and psychosomatic symptoms remained significantly more common than among healthy women.

Strengths and limitations

The strengths of this study include its nationwide, population-based design. Selection bias is possible, but is likely to be minimal because of our high response rates. Self-report bias is inherent in this study design, but our interviewers were highly trained eating disorder professionals used to diagnosing bulimia nervosa in both clinical and community settings.

Because this study focused on DSM-IV bulimia nervosa in the general population, the number of affected cases was relatively low despite the nationwide study setting. Our screening measures were based on self-report, and relied heavily on self-recognition of past and current symptoms of bulimia. Because of stigma and shame still commonly associated with reporting these symptoms, we may, despite our best efforts, have underestimated the true rates of bulimia nervosa. Our method of case detection, the SCID, may also have been less accurate in its ability to detect cases of bulimia nervosa, particularly subthreshold cases, than more extensive diagnostic interviews, such as the Eating Disorder Examination (Fairburn & Cooper, 1993). This may have influenced our ability to correctly identify unaffected co-twins. However, many other epidemiological studies have relied on similar methodology (Sullivan *et al.* 1998; Hudson *et al.* 2007) that has been shown to be reliable if multiple prompts are used (Wade *et al.* 2000). Our experienced diagnosticians were especially trained to use multiple prompts.

Because of the young age of our participants, outcomes beyond the 5-year clinical recovery rate were of an exploratory nature and limited power. Although relapses of bingeing and purging associated with bulimia nervosa are increasingly unusual after >1 year of abstinence (Field *et al.* 1997), we and others (Fairburn *et al.* 2000; Wade *et al.* 2006) found that residual symptoms were common after recovery. A better understanding of population outcomes of bulimia nervosa thus requires considerably longer periods of follow-up than was possible in our setting.

We could not directly assess the impact of treatment on outcome. We found that the detection of bulimia by the health-care system had no association with prognosis in this study. This result does not, however, demonstrate that treatment had no effect, because case detection does not necessarily imply proper treatment.

Finally, the aim of this study was not to study the familiarity and heritability of bulimia nervosa. Our two-stage sampling design was less than ideally suited for analyses of heritability, and heritability of bulimia nervosa and its component behaviors has been addressed extensively elsewhere (e.g. Bulik et al. 2007).

Conclusions

Although bulimia nervosa affects more than 2% of women during their lifetime, usually peaking at age 16–20 years, few women with bulimia nervosa are recognized in health-care settings. This is problematic, because bulimia nervosa is relatively long-standing and many clinically recovered women continue to experience residual symptoms over long periods of time after reaching abstinence from bingeing and purging.

Acknowledgments

We thank Professor Ian McKeague for statistical advice, Jutta Kaukoranta and Seija Salmelin for their help in conducting the interviews, and Eila Voipio, Kauko Heikkilä and Pia Ruokolinna for technical assistance. The design and conduct of this study were funded by NIAAA grants AA12502 and AA08315, Academy of Finland (44069 and 201461), and the European Union Fifth Framework Program (QLRT-1999-00916 and QL2-CT-2002-01254). Data preparation and analysis were supported by the Psychiatric Epidemiology Training Program at Columbia University (MH013043) and funds from the Ahokas, Jahnsson, Helsingin Sanomat, and Finnish Cultural Foundations (to A.K.-R.).

Declaration of Interest

None.

References

- Becker AE, Grinspoon SK, Klubanski A, Herzog DB (1999). Eating disorders. *New England Journal of Medicine* **340**, 1092–1098.
- Berkman ND, Lohr KN, Bulik CM (2007). Outcomes of eating disorders: a systematic review of the literature. *International Journal of Eating Disorders* **40**, 293–309.
- Bulik CM, Hebebrand J, Keski-Rahkonen A, Klump KL, Reichborn-Kjennerud T, Mazzeo SE, Wade TD (2007).

Genetic epidemiology, endophenotypes, and eating disorder classification. *International Journal of Eating Disorders* **40** (Suppl.), S52–S60.

- Cullberg J, Engström-Lindberg M (1988). Prevalence and incidence of eating disorders in a suburban area. *Acta Psychiatrica Scandinavica* **78**, 314–319.
- Curran L, Schmidt U, Treasure J, Jick H (2005). Time trends in eating disorder incidence. *British Journal of Psychiatry* **186**, 132–135.
- Fairburn CG, Cooper Z (1993). The Eating Disorder Examination, 12th edn. In *Binge Eating: Nature, Assessment and Treatment* (ed. C. G. Fairburn and G. T. Wilson), pp. 317–360. Guilford Press: New York.
- Fairburn CG, Cooper Z, Doll HA, 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.
- Fairburn CG, Harrison PJ (2003). Eating disorders. *Lancet* **361**, 407–416.
- Field AE, Herzog DB, Keller MB, West J, Nussbaum K, Colditz GA (1997). Distinguishing recovery from remission in a cohort of bulimic women: how should asymptomatic periods be described? *Journal of Clinical Epidemiology* **50**, 1339–1345.
- First MB, Spitzer RL, Gibbon M, Williams JBW (2003). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Nonpatient Edition*. Biometrics Res., New York State Psychiatric Institute: New York.
- Garfinkel PE, Lin E, Goering P, Speeg C, Goldbloom DS, Kennedy S, Kaplan AS, Woodside DB (1995). Bulimia nervosa in a Canadian community sample: prevalence and comparison of subgroups. *American Journal of Psychiatry* **152**, 1052–1058.
- Garner DM (1991). *Eating Disorder Inventory 2, Professional Manual*. Psychological Assessment Resources, Inc.: Odessa, FL.
- Hepworth N, Paxton SJ (2007). Pathways to help-seeking in bulimia nervosa and binge eating problems: a concept mapping approach. *International Journal of Eating Disorders* **40**, 493–504.
- Hoek HW, Bartelds AI, Bosveld JJ, van der Graaf Y, Limpens VE, Maiwald M, Spaaij CJ (1995). Impact of urbanization on detection rates of eating disorders. *American Journal of Psychiatry* **152**, 1272–1278.
- Hoek HW, van Hoeken D (2003). Review of the prevalence and incidence of eating disorders. *International Journal of Eating Disorders* **34**, 383–396.
- Hudson JI, Hiripi E, Pope Jr. HG, Kessler RC (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry* **61**, 348–358.
- Kaprio J, Pulkkinen L, Rose RJ (2002). Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Research* **5**, 366–371.
- Kaye WH, Greeno CG, Moss H, Fernstrom J, Fernstrom M, Lilienfeld LR, Weltzin TE, Mann JJ (1998). Alterations in serotonin activity and psychiatric symptoms after recovery from bulimia nervosa. *Archives of General Psychiatry* **55**, 927–935.

- Keel PK, Mitchell JE** (1997). Outcome in bulimia nervosa. *American Journal of Psychiatry* **154**, 313–321.
- Keski-Rahkonen A, Hoek HW, Susser ES, Linna MS, Sihvola E, Raevuori A, Bulik CM, Kaprio J, Rissanen A** (2007). Epidemiology and course of anorexia nervosa in the community. *American Journal of Psychiatry* **164**, 1259–1265.
- Keski-Rahkonen A, Sihvola E, Raevuori A, Kaukoranta J, Bulik CM, Hoek HW, Rissanen A, Kaprio J** (2006). Reliability of self-reported eating disorders: optimizing population screening. *International Journal of Eating Disorders* **39**, 754–762.
- Lehmann HE** (1983). Clinical evaluation and natural course of depression. *Journal of Clinical Psychiatry* **44**, 5–10.
- Lewinsohn PM, Striegel-Moore RH, Seeley JR** (2000). Epidemiology and natural course of eating disorders in young women from adolescence to young adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 1284–1292.
- Milos G, Spindler A, Schnyder U, Martz J, Hoek HW, Willi J** (2004). Incidence of severe anorexia nervosa in Switzerland: 40 years of development. *International Journal of Eating Disorders* **35**, 250–258.
- Mond J, Hay P, Rodgers B, Owen C** (2006). Self-recognition of disordered eating among women with bulimic-type eating disorders: a community-based study. *International Journal of Eating Disorders* **39**, 747–753.
- Rosenberg M** (1989). *Society and the Adolescent Self-Image. Revised edition*. Wesleyan University Press: Middletown, CT.
- Schousboe K, Willemsen G, Kyvik KO, Mortensen J, Boomsma DI, Cornes BK, Davis CJ, Fagnani C, Hjelmberg J, Kaprio J, De Lange M, Luciano M, Martin NG, Pedersen N, Pietiläinen KH, Rissanen A, Saarni S, Sorensen TI, Van Baal GC, Harris JR** (2003). Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Research* **6**, 409–421.
- Soundy TJ, Lucas AR, Suman VJ, Melton LJ** (1995). Bulimia nervosa in Rochester, Minnesota from 1980 to 1990. *Psychological Medicine* **25**, 1065–1071.
- Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA** (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press: Palo Alto, CA.
- Sullivan PF, Bulik CM, Kendler KS** (1998). The epidemiology and classification of bulimia nervosa. *Psychological Medicine* **28**, 599–610.
- Sutandar-Pinnock K, Woodside DB, Carter JC, Olmsted MP, Kaplan AS** (2003). Perfectionism in anorexia nervosa: a 6–24-month follow-up study. *International Journal of Eating Disorders* **33**, 225–229.
- Van Son GE, van Hoeken D, Bartelds AI, van Furth EF, Hoek HW** (2006). Time trends in the incidence of eating disorders: a primary care study in the Netherlands. *International Journal of Eating Disorders* **39**, 565–569.
- Wade TD, Bergin JL, Tiggemann M, Bulik CM, Fairburn CG** (2006). Prevalence and long-term course of lifetime eating disorders in an adult Australian twin cohort. *Australian and New Zealand Journal of Psychiatry* **40**, 121–128.
- Wade TD, Bulik CM, Kendler KS** (2000). Reliability of lifetime history of bulimia nervosa. Comparison with major depression. *British Journal of Psychiatry* **177**, 72–76.
- Walsh BT, Devlin MJ** (1998). Eating disorders: progress and problems. *Science* **280**, 1387–1390.
- Wilfley DE, Bishop ME, Wilson GT, Agras WS** (2007). Classification of eating disorders: toward DSM-V. *International Journal of Eating Disorders* **40** (Suppl. S), 123–129.
- Willi J, Giacometti G, Limacher B** (1990). Update on the epidemiology of anorexia nervosa in a defined region of Switzerland. *American Journal of Psychiatry* **147**, 1514–1517.
- Williams RL** (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics* **56**, 645–646.