Reliability of lifetime history of bulimia nervosa

Comparison with major depression

TRACEY D. WADE, CYNTHIA M. BULIK and KENNETH S. KENDLER

Background   Previous studies have found that the reliability of the lifetime prevalence of bulimia nervosa is low to moderate. However, the reasons for poor reliability remain unknown.

Aims   We investigated the ability of a range of variables to predict reliability, sensitivity, and specificity of reporting of both bulimia nervosa and major depression.

Method   Two interviews, approximately 5 years apart, were completed with 2163 women from the Virginia Twin Registry.

Results   After accounting for different base rates, bulimia nervosa was shown to be as reliably reported as major depression. Consistent with previous studies of major depression, improved reliability of bulimia nervosa reporting is associated with more severe bulimic symptomatology.

Conclusions   Frequent binge eating and the presence of salient behavioural markers such as vomiting and laxative misuse are associated with more reliable reporting of bulimia nervosa. In the absence of the use of fuller forms of assessment, brief interviews should utilise more than one prompt question, thus increasing the probability that memory of past disorders will be more successfully activated and accessed.

Declaration of interest   None.

Four studies have reported the reliability of diagnosis of bulimia nervosa over time. Interviews conducted within the same month have fair agreement (κ=0.42) (Bushnell *et al*, 1990). A 10-year follow-up also found moderate agreement for some behaviours but not others: binge eating (κ=0.47); self-induced vomiting (κ=0.49); laxative misuse (κ=0.50); diet pills (κ=0.45); and fasting (κ=0.25) (Field *et al*, 1996). Use of a fuller assessment on at least one occasion seems to promote moderate agreement (κ=0.59) (Wade *et al*, 1997). Reliability of a more broadly defined phenotype of bulimia nervosa may produce lower agreement (κ=0.28) (Bulik *et al*, 1998). This study was designed to further explore predictors of reliability of the lifetime diagnosis of bulimia nervosa in comparison with predictors of reliability of a lifetime diagnosis of major depression, which were assessed with the same diagnostic instrument in a large population-based sample of female twins.

METHOD

Participants

The data for this report are from a population-based longitudinal study of Caucasian female twins drawn from the Virginia Twin Registry (VTR). The VTR was formed from a systematic review of all birth records in the Commonwealth of Virginia (USA) after 1918. Twins were eligible to participate if they were born between 1934 and 1971 and both members had previously responded to a mailed questionnaire completed over 1987–88 (individual response rate of 64%). Data used in the present study are from the first interview wave and accompanying self-report personality measures, and the third interview wave, which will be called Time 1 interview and Time 2 interview, respectively. At Time 1 (1987–89), 92% of the eligible individuals (*n*=2163) were interviewed (90% face to face and the remainder by telephone). The mean age of the twins was 29.3 years (s.d.=7.7, range 17–54 years). Time 2 (1991–93) occurred on average 5.1 years (s.d.=0.4) later. Written informed consent was obtained prior to face-to-face interviews and verbal assent prior to telephone interviews.

Measures

Interviews were conducted blind to information about co-twins. Information about interviewer characteristics has been presented elsewhere (Kendler *et al*, 1991). A narrow definition of lifetime bulimia nervosa, or one that conformed strictly to DSM–III–R (American Psychiatric Association, 1987) criteria, was used. In addition, in order to maximise statistical power in the study of a low prevalence disorder, a broad definition of lifetime bulimia nervosa was adopted where the DSM–III–R 'D' criterion was omitted because there appear to be few meaningful differences between women who binge and use associated weight-loss methods twice a week and those who engage in such behaviours less than twice a week (Garfinkel *et al*, 1995; Sullivan *et al*, 1998). This broad category differs slightly from its previous use (Bulik *et al*, 1998), in that it includes women with a wider range of concern about their body shape and weight, from "a lot more concerned than most women your age" to "a little bit more concerned".

At the first interview there was one probe question ("Have you ever in your life had eating binges during which you ate a lot of food in a short period of time?"). If this was answered negatively, no further questions were asked. At the second interview, a further probe question was asked, relating to weight loss behaviours ("Have you ever made yourself throw up as a means of controlling your shape and weight?"). If these were both answered negatively, no further questions were asked.

The diagnosis of DSM–III–R major depression was made using questions from the Structured Clinical Interview for DSM–III–R (SCID) (Spitzer *et al*, 1992). Numerous probe questions were used to ascertain the presence of depressive symptoms. Initially, occurrence of major depression over the past year was assessed, using a probe question for each one of the diagnostic criteria. Then major depression over the lifetime (excluding the past year) was assessed with two probe questions.

A description of the variables examined for predictive value of reliability is provided

in Table 1: all predictor variables are from the Time 1 interview period unless otherwise noted.

Statistical analyses

Agreement between Time 1 and Time 2 diagnoses were examined using the kappa coefficient (κ), tetrachoric correlations, and the Yule's Y statistic. Yule's Y is less dependent on the base rate than κ, which permits a more direct comparison between the higher prevalence major depression and the lower prevalence bulimia nervosa. We also calculated sensitivity – the proportion of true cases correctly identified (risk for false-negatives) – and specificity – the proportion of true non-cases correctly identified (risk for false-positives). For the purpose of these calculations, the Time 2 assessment was chosen as the standard, as it contained more probe questions than the Time 1 assessment. One would expect sensitivity to be lower for more prevalent disorders and specificity to be higher for less prevalent disorders.

The ability of variables to predict reliability, sensitivity and specificity was then examined using logistic regression. Results are presented as odds ratios with 95% CIs. As twin pair observations are correlated, the assumption of independent sampling is violated, and we therefore used generalised estimating equation (GEE) modelling (Zeger *et al*, 1988) to adjust standard errors for non-independent observations using the GENMOD procedure.

Finally, separate stepwise logistic regressions were used to examine the relative importance of the significant predictors for reliability, sensitivity and specificity of reporting bulimia nervosa. All analyses were carried out with SAS version 7.0 (SAS Institute, 1996).

RESULTS

Agreement between interviews

For the purposes of this study, women who reported lifetime bulimia nervosa at Time 2 but not Time 1, and who reported age of onset as being after Time 1, were considered to have developed bulimia nervosa between the two assessments. These women were removed from further analysis so that onsets that occurred between Time 1 and Time 2 would not be confounded with unreliable recall. This resulted in two women being removed when considering narrowly defined bulimia nervosa and 11 women being removed when considering broadly defined bulimia nervosa. Reliability of major depression has previously been considered with complete twin pairs only (Foley *et al*, 1998). As completeness of twin pairs was irrelevant to our analyses, we considered data from all twins, thus increasing the number of women studied and producing a slightly higher κ value than previously reported. When onset of first-episode major depression between Times 1 and 2 was considered, 113 women were removed from further analysis. Taking into account the lower base rate dependent measures (tetrachoric correlations and Yule's Y), narrowly defined bulimia nervosa is the most reliable diagnosis, and the reliability of broadly defined bulimia nervosa and major depression are similar (Table 2). The bulimia nervosa diagnoses have the lowest risk for assigning false-positive cases, but the highest risk for assigning false-negative cases.

Clinical features predicting reliability of bulimia nervosa

By far the largest category of women with unreliably reported bulimia nervosa included women who met the full criteria at one interview and gave negative replies to the probe question/s at the other interview – for narrowly defined bulimia nervosa this occurred approximately one-third of the time, for broadly defined bulimia nervosa it occurred approximately half the time. Reported use of self-induced vomiting or laxative misuse at either interview significantly predicted reliability ($P=0.005$, odds ratio=3.48, 95% CI 1.45–8.35). The likelihood of reporting the behaviour associated with bulimia nervosa at Time 2 was dependent on the type of behaviour reported at Time 1 (see Table 3). The most memorable weight loss behaviour was self-induced vomiting (with the odds of reporting vomiting at the second interview 34 times higher if vomiting was reported at the first interview) and laxative misuse (with the odds of reporting laxative abuse at the second interview 28 times higher if laxative abuse was reported at first interview). In contrast,

Table 1 Predictor variables used to examine influences on the reliability of reporting of lifetime bulimia nervosa and major depression (predictor variables are from the Time 1 interview period unless otherwise noted)

Variable type	Description of variable
Demographics	Years of education (highest grade of school or year of college completed), annual salary, parental education, urbanisation
	Population size of area lived in at time of interview
	Annual salary
	Parental education
Age	Age at first interview
	Age of developing bulimia nervosa recalled at both Time 1 and Time 2
Current symptomatology	20 items of the Padua Inventory (Sanavio, 1988) assessing levels of obsessive and compulsive symptomatology
	Total score of the Symptom Check-List (Derogatis, 1975), assessing depression, panic and agoraphobia, somatisation and anxiety, and sleep difficulties (Kendler *et al*, 1994) from the previous 30 days
Lifetime comorbidity	Structured Clinical Interview (Spitzer *et al*, 1992) for:
	DSM–III–R alcohol dependence and panic disorder
	DSM–III phobias
	DSM–III–R symptoms and DSM–III-duration generalised anxiety disorder
Personality and attitudes	Extroversion and neuroticism (Eysenck & Eysenck, 1975)
	Altruism (7 items from the Interpersonal Reactivity Index; Davis, 1980)
	Interpersonal dependency (Hirschfeld *et al*, 1977)
	Mastery (reversed coding of the powerlessness sub-scale of the Alienation subtest; Maddi *et al*, 1979)
	Dispositional optimism (Scheier & Carner, 1985)
	Self-esteem (Rosenberg, 1965)
	Locus of control (resourcefulness sub-scale of the Attributional Style Questionnaire; Peterson *et al*, 1982)

odds of recalling strict dieting or fasting at Time 2 were only about twice as high if such behaviour was reported at Time 1. Binge eating was less likely to be recalled than either self-induced vomiting or laxatives, but more likely to be remembered than other weight loss behaviours.

The more detailed Time 2 data were used to investigate any differences in frequency of eating disorder behaviours between those women with reliably reported bulimia nervosa and those women with unreliably reported bulimia nervosa. Results are summarised in Table 4. The strongest association exists between reliability and frequency of binge eating. For both narrowly defined and broadly defined bulimia nervosa, a higher monthly frequency of binge eating predicted more reliable reporting.

Predictors of reliability, sensitivity and specificity of bulimia nervosa and major depression

For the remaining analyses, there was insufficient power to calculate the odds ratio for narrowly defined bulimia nervosa. Therefore, only results for broadly defined bulimia nervosa and major depression are reported here.

Reliability

For bulimia nervosa, more years of education, parental education and decreased likelihood of lifetime major depression were significantly associated with more reliable reporting (data not shown). The women with reliably reported major depression were significantly older than the women with unreliably reported major depression, had higher levels of obsessiveness, general anxiety and depression, and were more likely to experience lifetime generalised anxiety disorder (GAD), panic disorder and simple phobias. There was also considerable influence of personality on the reliability of major depression reporting, where women who reliably reported major depression were significantly more dependent, experienced less mastery, were less optimistic, had lower self-esteem and were more neurotic. In other words, this group appeared to be generally more impaired.

Sensitivity

Increased ability to detect true cases of bulimia nervosa was predicted by more years of parental education and lower levels of altruism. However, because of the lack of convergence occurring in the logistic regression and the consequent inability to produce odds ratios, not all variables could be satisfactorily examined. Increased sensitivity of major depression was predicted by a lower financial status, higher levels of obsessive symptomatology and neuroticism, increased risk for lifetime comorbidity, especially GAD, and lower levels of mastery and optimism.

Table 2 Agreement between Time 1 and Time 2 diagnosis – lifetime bulimia nervosa, both narrowly and broadly defined, and lifetime major depression

Time 1 assessment	Time 2 assessment	
	Bulimia nervosa	No bulimia nervosa
Narrowly defined bulimia nervosa[1]		
Bulimia nervosa	11	31
No bulimia nervosa	8	1845
Broadly defined bulimia nervosa[2]		
Bulimia nervosa	33	49
No bulimia nervosa	59	1745
	Major depression	No major depression
Major depression[3]		
Major depression	387	277
No major depression	130	988

1. $\kappa=0.35$, tetrachoric $r=0.81$, Yule's $Y=0.80$, sensitivity$=0.57$, specificity$=0.98$.
2. $\kappa=0.35$, tetrachoric $r=0.70$, Yule's $Y=0.63$, sensitivity$=0.36$, specificity$=0.97$.
3. $\kappa=0.49$, tetrachoric $r=0.72$, Yule's $Y=0.53$, sensitivity$=0.75$, specificity$=0.78$.

Table 3 Reliability of reporting of behaviours associated with bulimia nervosa at Time 2, given they were reported at Time 1

Weight loss behaviour	κ	Odds ratio (95% CI)	P
Self-induced vomiting	0.56	34.11 (15.74–73.94)	<0.0001
Laxative misuse	0.50	28.74 (9.41–87.81)	<0.0001
Binge eating	0.37	6.13 (4.84–7.78)	<0.0001
Exercise	0.23	3.61 (1.77–7.38)	0.0003
Strict dieting	0.15	2.29 (1.12–4.67)	0.02
Fasting	0.13	2.41 (0.97–5.98)	0.05

Table 4 The impact of symptom severity on reliability of diagnosis of lifetime bulimia nervosa

Symptom	Narrowly-defined bulimia nervosa		Broadly-defined bulimia nervosa		
	Mean reliable	Mean unreliable	Mean reliable	Mean unreliable	Odds ratio (95% CI)
Frequency of binges each month	26.73	9.45	12.21	5.64	1.07 (1.02–1.12)**
Duration of binge eating (in weeks)	248.40	202.90	215.46	195.21	1.00 (0.99–1.01)
Frequency of vomiting each month	59.78	20.60	39.28	7.67	1.04 (0.97–1.12)
Duration of vomiting (in weeks)	145.78	58.73	91.67	32.04	1.01 (0.99–1.03)

1. Insufficient numbers for convergence to produce odds ratio.
**$P<0.01$.

Specificity

Increased ability to correctly identify true non-cases of bulimia nervosa was predicted by lower levels of current symptomatology, decreased risk for lifetime comorbidity, higher levels of mastery and self-esteem and lower neuroticism. Increased specificity of major depression was predicted by a higher financial status, lower levels of current symptomatology, decreased risk of lifetime comorbidity, lower levels of altruism, dependency and neuroticism and greater optimism.

Multivariate contribution of predictor variables to reliability, sensitivity and specificity

The relative contributions of those predictor variables shown to significantly predict reliability of reporting of broadly defined bulimia nervosa were examined in a stepwise regression model, including reported use of either self-induced vomiting or laxatives, frequency of binge eating, years of education, educational status of parents and presence of lifetime major depression. The variables retained in the equation that predicted more reliable reporting of bulimia nervosa were decreased likelihood of lifetime major depression at either Time 1 or Time 2 ($\chi^2=5.18$, $P=0.02$), use of either self-induced vomiting or laxatives ($\chi^2=4.84$, $P=0.03$), and greater frequency of binges each month ($\chi^2=4.28$, $P=0.04$). Predictors of greater reliability of major depression reporting (including only those significant predictor variables) included greater likelihood of GAD ($\chi^2=23.17$, $P<0.0001$), a higher score on the Symptom Check-List (Derogatis, 1975) at Time 2 ($\chi^2=7.28$, $P=0.007$), and increased obsessionality ($\chi^2=4.83$, $P=0.03$).

Due to the low predictive power of the sensitivity measure, this was not examined in a multiple regression for bulimia nervosa. Of those variables that significantly predicted greater sensitivity for major depression in the univariate analyses, two were retained in the final equation, including greater likelihood of lifetime GAD ($\chi^2=28.92$, $P<0.0001$) and lower financial status ($\chi^2=7.03$, $P=0.008$). Of those variables that significantly predicted greater specificity of bulimia nervosa in the univariate analyses, the following were retained in the final equation: decreased likelihood of lifetime major depression ($\chi^2=10.37$, $P=0.001$) and panic disorder ($\chi^2=5.88$, $P=0.02$), and increased levels of mastery ($\chi^2=6.64$, $P=0.01$). Correspondingly, variables that best predicted major depression specificity were a lower likelihood of lifetime GAD ($\chi^2=92.22$, $P<0.0001$) and alcohol dependency ($\chi^2=16.91$, $P<0.0001$) and lower levels of altruism ($\chi^2=5.37$, $P=0.02$).

DISCUSSION

From previous literature, using a base rate sensitive measure (κ), bulimia nervosa would appear to be a less reliable diagnosis than major depression, usually showing low to modest agreement between assessments (Bushnell *et al*, 1990; Field *et al*, 1996; Bulik *et al*, 1998). We replicate the finding that base rate sensitive measures (i.e. κ) show bulimia nervosa to be less reliably diagnosed than major depression.

However, given the much greater prevalence of major depression than bulimia nervosa, use of measures less dependent on the base rate may be a more appropriate way of comparing reliabilities. The use of such measures (i.e. Yule's Y) shows bulimia nervosa to be as reliably diagnosed as major depression. As can be predicted, it is more difficult to label a true non-case of bulimia nervosa as a case than it is major depression. The fairly unique behavioural markers for bulimia nervosa (e.g. binge eating, vomiting) compared to the less discrete features of major depression, which can be shared with other disorders (e.g. insomnia, fatigue, diminished ability to concentrate), may amplify this effect. On the other hand, it is much more difficult to accurately identify true cases of bulimia nervosa than major depression. The occurrence of past major depression may be more accessible to memory as the symptoms are more likely to be reminiscent of aspects of current life experience than are those of past bulimia nervosa. In addition, the presence of more probe questions in the interview for major depression than bulimia nervosa may account for the greater difficulty in detecting bulimia nervosa cases than major depression cases. This suggestion is consistent with the body of neuropsychological literature, which shows that verbal prompts improve verbal recall for both younger and older adults (Cherry *et al*, 1996).

Salience of behavioural markers

In terms of overall reliability, we replicated the findings of Field *et al* (1996) where the majority of unreliable cases were women who reported full symptoms of bulimia nervosa on one occasion and responded negatively to probe questions on the other. Of all the behaviours associated with bulimia nervosa reported at Time 1, it was the presence of self-induced vomiting and laxative misuse that were most likely to be remembered at Time 2. This suggests vomiting and laxatives are more salient behavioural markers than other weight loss behaviours, and thereby less vulnerable to memory decay. However, a higher monthly frequency of binge eating rather than any weight loss behaviour significantly predicts reliable reporting of lifetime bulimia nervosa. As not all women use vomiting or laxatives, the frequencies of these behaviours may have had insufficient predictive power. These findings concur with studies on the reliability of major depression, which suggest the more severe the symptomatology, the more memorable the disorder (Aneshensel *et al*, 1987; Foley *et al*, 1998).

Role of sensitivity and specificity in determining reliability

There appear to be more differences than similarities in the profiles of overall predictive reliability of bulimia nervosa and major depression. Reliability of major depression reporting appears to be affected by overall level of functioning of the individual. The less well the person, as indicated by a number of measures including personality, current symptomatology and lifetime psychopathology, the more likely they were to reliably recall having had major depression. In contrast, there was no effect of personality or attitudes on reliability of bulimia nervosa reporting, and the strongest predictor, apart from the behavioural markers, was a lower likelihood of lifetime major depression. This finding can be explained by examination of sensitivity and specificity. The presence of true cases of major depression is marked by increased problems with psychiatric and personality functioning (unfortunately our ability to detect true cases of bulimia nervosa was limited). Conversely, the detection of true non-cases of both bulimia nervosa and major depression was marked by fewer problems with psychiatric and personality functioning. This would suggest that the overall reliability of bulimia nervosa seems to be characterised more by its ability to accurately detect non-cases, whereas the overall reliability of major depression is characterised more by its ability to detect cases.

A simple comparison of general reliability measures across psychiatric diagnoses is insufficient to elucidate the nature and mechanisms of unreliability. A more useful approach is to examine specific aspects of reliability of reporting, such as sensitivity and specificity. Given that the majority of population-based epidemiological studies utilise structured clinical interviews to identify cases of bulimia nervosa similar to the ones used in this investigation, several strategies can be employed to improve reliability of reporting in the context of such interviews. Incorporating more than one occasion of measurement (Bulik *et al*, 1998) and using more specialised assessment instruments (Wade *et al*, 1997) can improve reliability. In addition, the inclusion of a greater number of probe questions can increase the probability that memory of past disorders will more successfully be activated and accessed, thus increasing the detection of true cases.

CLINICAL IMPLICATIONS

■ Individuals who self-induce vomiting or misuse laxatives have more reliable recall of bulimia nervosa.

■ More frequent occasions of binge eating are associated with more reliable reporting of the disorder.

■ It is wise to include several probe questions about several behaviours when assessing bulimia nervosa (e.g. binge eating, vomiting, laxative misuse) rather than using a single probe to stimulate recall.

LIMITATIONS

■ Only women were assessed, therefore these results cannot be applied to men.

■ We have low power to draw any conclusions about variables that predict sensitivity of bulimia nervosa.

■ The use of Time 2 assessment as the standard is based on this assessment having only one more probe question than Time 1 assessment.

TRACEY D. WADE, PhD, CYNTHIA M. BULIK, PhD, KENNETH S. KENDLER, MD, Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia, USA

Correspondence: Associate Professor Cynthia M. Bulik, Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University, PO Box 980126, Richmond, VA 23298−0126, USA. Tel: 001 804 828 8133; Fax: 001 804 828 1471

REFERENCES

American Psychiatric Association (1987) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM−III−R). Washington, DC: APA.

Aneshensel, C. S., Estrada, A. L., Hansell, M., et al (1987) Social psychological aspects of reporting behaviour: Lifetime depressive episode reports. *Journal of Health and Social Behaviour*, **28**, 232−246.

Bulik, C. M., Sullivan, P. F. & Kendler, K. S. (1998) Heritability of binge-eating and broadly defined bulimia nervosa. *Biological Psychiatry*, **44**, 1210−1218.

Bushnell, J. A., Wells, J. E., Hornblow, A. R., et al (1990) Prevalence of three bulimia syndromes in the general population. *Psychological Medicine*, **20**, 671−680.

Cherry, K. E., Park, D. C., Frieske, D. A., et al (1996) Verbal and pictorial elaborations enhance memory in younger and older adults. *Aging, Neuropsychology, and Cognition*, **3**, 15−29.

Davis, M. H. (1980) A multi-dimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology*, **10**, 85.

Derogatis, L. R. (1975) *Symptom Check-List−90−Revised*. Minneapolis, MN: National Computer Systems Inc.

Eysenck, H. J. & Eysenck, S. B. G. (1975) *Manual of the Eysenck Personality Questionnaire*, London: Hodder & Stoughton.

Field, A. E., Colditz, G. A., Herzog, D. B., et al (1996) Disordered eating: Can women accurately recall their binging and purging behaviors 10 years later? *Obesity Research*, **4**, 153−159.

Foley, D. L., Neale, M. C. & Kendler, K. S. (1998) Reliability of lifetime history of major depression: implications for heritability and co-morbidity. *Psychological Medicine*, **28**, 857−870.

Garfinkel, P. E., Lin, E., Goering, P., et al (1995) Bulimia nervosa in a Canadian community sample: Prevalence and comparison of subgroups. *American Journal of Psychiatry*, **152**, 1052−1058.

Hirschfeld, R. M. (1977) A measure of interpersonal dependency. *Journal of Personality Assessment*, **41**, 610−618.

Kendler, K. S., MacClean, C., Neale, M. C., et al (1991) The genetic epidemiology of bulimia nervosa. *American Journal of Psychiatry*, **148**, 1627−1637.

___ , Walters, E. E., Truett, K. R., et al (1994) Sources of individual differences in depressive symptoms: Analysis of two samples of twins and their families. *American Journal of Psychiatry*, **151**, 1605−1614.

Maddi, S. R., Kobasa, S. C. & Hoover, M. (1979) An alienation test. *Journal of Humanist Psychology*, **19**, 73−76.

Peterson, C., Semmel, A., vonVaeyer, C., et al (1982) The attributional style questionnaire. *Cognitive Therapy Research*, **6**, 287−299.

Rosenberg, M. (1965) *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

SAS Institute Inc. (1996) *SAS/STAT=AE Software: Changes and Enhancements. Release 6.11.* Cary, NC: SAS Institute Inc.

Sanavio, E. (1988) Obsessions and compulsions: the Padua Inventory. *Behaviour Research and Therapy*, **26**, 169−177.

Scheier, M. G. & Carner, C. S. (1985) Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health Psychology*, **4**, 219−247.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., et al (1992) The structured clinical interview for DSM−III−R (SCID) I: history, rationale and description. *Archives of General Psychiatry*, **49**, 624−629.

Sullivan, P. F., Bulik, C. M. & Kendler, K. S. (1998) The epidemiology and classification of bulimia nervosa. *Psychological Medicine*, **28**, 599−610.

Wade, T., Tiggemann, M., Martin, N. G., et al (1997) A comparison of the Eating Disorder Examination and a general psychiatric interview. *Australian and New Zealand Journal of Psychiatry*, **31**, 852−857.

Zeger, S. L., Liang, K. Y. & Albert, P. S. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049−1060.