

Key words: eating disorders, eating disorders not otherwise specified, screening, false negatives, Eating Attitudes Test (EAT40), Eating Disorder Inventory (EDI).

Screening for eating disorders: false negatives and eating disorders not otherwise specified

Giuseppe Vetrone, MD*
Massimo Cuzzolaro, MD**
Italo Antonozzi, MD***
Paul E Garfinkel, MD****

* University of Rome Tor Vergata,
Department of Philosophical Researches

** University of Rome La Sapienza,
Department of Child and Adolescent
Psychiatry

*** University of Rome La Sapienza,
Department of Experimental Medicine

**** University of Toronto

ITALIA, CANADA

ABSTRACT – *Objective:* To study the problem of false negatives in the screening for eating disorders.

Subjects and Methods: We administered the Eating Attitudes Test (EAT40) and the Eating Disorder Inventory (EDI) to 186 Italian female students (age 17.68 ± 0.9 ; BMI 20.84 ± 2.69). Then we submitted not only high-scorers but also all the subjects to a semi-structured diagnostic interview (Eating Disorder Examination - EDE 12.0D). The diagnosis of eating disorders not otherwise specified (EDNOS) was applied to cases which met all the DSM-IV criteria for AN and/or BN with one exception.

Results: 17 girls (9.1%) fulfilled diagnostic criteria for eating disorders: 1 bulimia nervosa full-syndrome and 16 EDNOS. Of the 17 girls 11 were EAT low-scorers (< 30) and 8 were EDI low-scorers (< 50); 5 subjects scored below the cutoff on both instruments. We calculated sensitivity (35.3%), specificity (88.8%), positive predictive value (PPV, 24.0%) and negative predictive value (NPV, 93.2%) of the EAT40; the respective values for the EDI were 52.9% (sensitivity), 85.2% (specificity), 26.4% (PPV) and 94.7% (NPV).

Discussion and conclusion: Our data show that the introduction of the EDNOS diagnoses increases the PPV of the two questionnaires but lowers their sensitivity. We conclude that using a two stage screening approach leads to a very high rate of false negatives with a significant underestimation of the prevalence of eating disorders, particularly of EDNOS.

Introduction

Detecting cases of eating disorders (ED) among large populations is of interest to clinicians and health planners, as well as epidemiologists. Most ED screenings carried out on community samples use a two-stage method in order to limit costs: in the first stage, a self-report questionnaire is administered; in the second stage a clinical interview is undertaken only in subjects thought to be possible cases, on the basis of a predefined scoring on the questionnaire (cutoff point).

In 1998 Schoemaker wrote that the problem of false negatives remained usually unappreciated (Schoemaker 1998). Therefore we consulted the Medline database (1999-2005) and we found very little research on eating disorders in which the problems of sensitivity of screening tests and especially of false negatives were analysed. It has left a concern about the frequency of false negatives, due too to the tendency of some subjects with ED to deny or hide their illness. Although this does not modify the negative predictive value (NPV) significantly (because of low prevalence of ED), it might be such as to produce a systematic underestimation of the prevalence of these disorders. If this were true the effectiveness of the screening studies would diminish significantly (Schoemaker 1998).

To verify this hypothesis, we decided to carry out an epidemiological study on a community sample in which all subjects were examined in the first stage with commonly used, well validated self-administered questionnaires. In the second stage a structured interview was undertaken. What distinguished this study was that the diagnostic interview was conducted not only with possible cases—selected on the basis of the questionnaire scores—but also with all subjects of

the original sample, including the low scorers. The diagnoses taken into account in the study were those in the DSM-IV (American Psychiatric Association 1994): Anorexia Nervosa (AN) full syndrome, Bulimia Nervosa (BN) full syndrome and Eating Disorders Not Otherwise Specified (EDNOS).

Subjects and methods

Subjects

Participants included the female population of the 4th year of two secondary state schools in the city of Rome, Italy. All the girls who were absent on the day of the testing were examined a few days later, giving a final sample size of 186 subjects. This sample size was the best compromise between the requirements and aims of the study and the costs due to the administration of a semi-structured clinical interview to all subjects.

Procedure and instruments

A two-stage method was followed. First, two questionnaires were administered: Eating Attitudes Test—EAT40 (Garner & Garfinkel 1979, Cuzzolaro & Petrilli 1988) and Eating Disorder Inventory—EDI (Garner *et al.* 1983). We used the original version of EDI (64 items) rather than EDI2 (91 items) (Garner 1991) to reduce the total length of the questionnaires. In this phase, weight and height were also measured and the BMI was calculated. Second, all the subjects, including the low scorers, were submitted to a detailed semi-structured clinical interview to establish a definite diagnosis, Eating Disorder Examination-EDE 12.0D (Cooper *et al.* 1987, Fairburn & Cooper 1993, Ricca *et al.* 1996) used as a gold standard for diagnosis. The inter-

views were performed, after a period of training, by psychiatrists and psychologists with at least two years of clinical experience with eating disordered patients. Diagnostic doubts were discussed jointly with a psychiatrist with a great deal of experience in the field of ED. The diagnosis of partial syndromes (EDNOS) (Crow *et al.* 2002, Norring & Palmer 2005) was applied to cases which met all the DSM-IV criteria for AN and/or BN with one exception.

Statistical Methods

We calculated sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of EAT40 and EDI; a ROC analysis (Receiver Operating Characteristics) was also carried out in order to evaluate discriminatory ability of the questionnaires.

Results

The mean age of the final sample of 186 girls was 17.68 (SD 0.9; range 16-19) with a mean BMI (body mass index) value 20.84 kg/m² (SD 2.69). The mean EAT40 total score was 16.61 (SD 11.49); the mean EDI total score was 32.92 (SD 21.27). The scores on these scales are quite typical of females in late adolescence and living in

Western countries (Garner & Garfinkel 1979, Cuzzolaro & Petrilli 1988, Garfinkel & Newman 2001, Garner *et al.* 1983, Canals *et al.* 2002, Rodriguez-Cano *et al.* 2005).

Using the EDE-12.0D 17 subjects were found to have an ED: 1 (0.5%) BN full-syndrome and 16 (8.6%) EDNOS (3 AN partial syndrome and 13 BN partial syndrome).

Table I shows data concerning sensitivity and specificity, PPV and NPV of EAT40 and EDI, obtained by using the customary cutoff points: 30 for EAT and 50 for EDI (total scores). Of the 17 girls who fulfilled diagnostic criteria for eating disorders 11 were EAT low-scorers (< 30, EAT false negatives) and 8 were EDI low-scorers (< 50, EDI false negatives). Five subjects scored below the cut-off on both instruments (EAT and EDI false negatives). Furthermore, we calculated sensitivity, specificity, PPV and NPV of the Drive for Thinness (DT) scale, choosing value 14 as cutoff point (Garner *et al.* 1983, Garner 1991).

Table II provides data concerning sensitivity, specificity, PPV and NPV of the two questionnaires, choosing, on the basis of the ROC analysis, two different cutoff points for each questionnaire: the total scores which ensured the most balanced sensitivity/specificity ratio were 19 for EAT40 and 37 for EDI; the total scores which ensured a sensitivity > 90% were 16 for EAT40 (specificity 65.1%) and 20 for EDI (specificity 32.0%).

Table I
Sensitivity, specificity, PPV, NPV of EAT40 (cutoff 30) and EDI (cutoff 50).

	EAT40 total score	EDI total score	EDI Drive for Thinness score
Cutoff point	≥ 30	≥ 50	≥ 14
High scorers (≤ cutoff point)	25 (13.4%)	34 (18.3%)	22 (11.8%)
Sensitivity	35.3%	52.9%	23.5%
Specificity	88.8%	85.2%	89.9%
PPV	24.0%	26.4%	19.1%
NPV	93.2%	94.7%	91.2%

Table II
Sensitivity, specificity, PPV, NPV of the EAT40 (cutoff 19 or 16) and EDI (cutoff 37 or 20).

	EAT40 total score	EDI total score
Cutoff point	19	37
Sensitivity	70.6%	70.6%
Specificity	70.4%	69.2%
PPV	19.4%	18.2%
NPV	95.9%	95.8%
Cutoff point	16	20
Sensitivity	94.1%	94.1%
Specificity	65.1%	32.0%
PPV	21.3%	12.2%
NPV	99.9%	98.2%

Table III displays the confirmatory data concerning the areas under the ROC curves of the questionnaires (a random classifier

has an area of 0.5; a classifier with an ideal discriminatory ability has an area of 1.0) and Figure 1 shows the ROC curves.

Table III
ROC analysis (Receiver Operating Characteristics).

	EAT40	EDI
Area under the ROC curves *	0.80	0.74
95% confidence interval of area	0.72 - 0.88	0.60 - 0.87

* The difference among the two areas is not statistically significant.

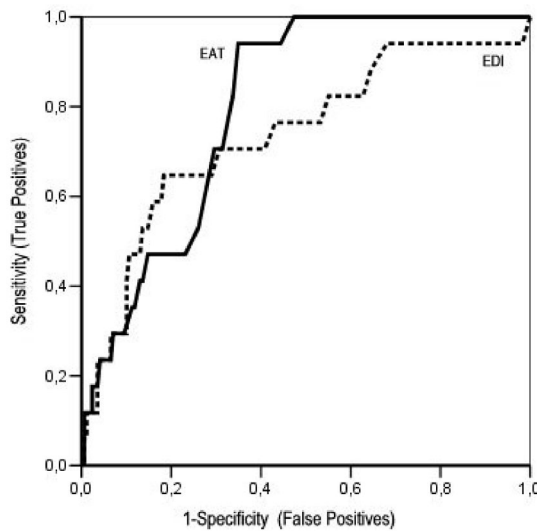


Figure 1. ROC curves for both EAT and EDI.

Discussion

ED screenings carried out on community samples have used, for the most part, a two-stage design. This type of methodology has considerable value but, at least, four important limits:

a. Use of any self-report can be affected by denial and social desirability.

b. Not all people with AN have high levels of fat phobia and drive for thinness. Strober *et al.* (1999) and Ramacciotti *et al.* (2002) found that about 15-20% of cases are in this category. This group have lower scores on other measures of eating disorder psychopathology (e.g., the EDI) as well. In Asian samples the frequency of the non-fat-phobic state was even higher (35%) (Lee *et al.* 2001) than in North America and Western Europe.

c. These instruments assess the individual at one point in time. Symptoms of the eating disorders can wax and wane over time (Fairburn & Harrison 2003). In a recent study (Lee *et al.* 2001), fat-phobic anorexics often changed over time; about 31% switched and endorsed non-fat-phobic items. In other words, these attributions are not static but they vary according to environment, life span and body weight.

d. Because the eating disorders are relatively uncommon in a population even when using questionnaires with adequate specificity, the PPV is low. Many of those interviewed in the second stage prove to be false positives and this increases the costs, thus reducing the efficiency of the screening studies.

Specificity and PPV of a test used can be calculated, while sensitivity, frequency of false negatives and NPV of the same tests are usually not available. This is because “a picture of the negative predictive value of a

screening test can only be achieved when all the subjects, including the low scorers, are interviewed in the second stage” (Beglin & Fairburn 1992). Such studies have not been performed to determine these estimates for eating disorders (Schoemaker 1998). Furthermore, the evaluation of sensitivity and specificity of the questionnaires are often based on the validation studies performed on clinical samples with ED full syndromes compared with control groups, not infrequently with a very small sample size.

To reduce the number of false negatives in the selection of the possible cases many researchers have added other criteria (e.g. BMI, amenorrhea) to the scores of the self-administered scales, but only very few researchers have tried to detect the false negatives by a direct diagnostic assessment.

For example, a recent Spanish two-stage survey tackled the problem of false negatives in the epidemiology of ED (Rodríguez-Cano *et al.* 2005). At the second stage of this research, 159 probable cases (on the basis of the self-administered questionnaires) and 150 controls (randomly selected from the remaining sample) were examined by means of a structured interview, the Spanish version of the SCAN system or Schedules for Clinical Assessment in Neuropsychiatry (Vazquez-Barquero *et al.* 1994). The authors found that 4 subjects, all women (3.6% of the 112 females of the control group) were false negatives (*both* EAT and BITE low scorers): the SCAN system identified 3 cases of EDNOS and 1 of AN full syndrome. If we compare these data with those of our research we can find similar figures. The false negatives (*both* EAT and BITE low scorers) were 3.6% in the Spanish study; our false negatives (*both* EAT and EDI low scorers) were 2.7%. T. Rodríguez-Cano and her colleagues conclude that the relatively high percentage of

false negatives, for the most part EDNOS, indicate that up to now the prevalence of ED in Spanish adolescents could be underestimated and underline the importance of carrying out epidemiological studies which would include standardized interview controlling possible false negatives.

What about EAT and EDI? The results of our study show that if we use the cutoff points suggested by the authors (30 for EAT40 and 50 for EDI) the sensitivity of the two questionnaires is very low. The figures are clearly inferior to those mentioned in the original validation studies (Garner & Garfinkel 1979, Garner *et al.* 1983). However, we must underscore that those studies were conducted only with patients affected by AN full syndrome (Garner & Garfinkel 1979, Garner *et al.* 1983). In our research, even by selecting the cutoff points which ensure the most balanced sensitivity/specificity ratio, the two questionnaires recognise only about 70% of the cases. On the other side, lower cutoff points with a sensitivity > 90% produce a too high number of false positives (59 for EAT and 115 for EDI in a total sample of 186 subjects). To sum up, even though the areas under the ROC curves indicate certain discriminant ability, the percentages are not adequate for accurate case identification.

We think that this low sensitivity might be explained, in the first place, by the fact that all cases in our sample, except one case of BN, which showed up in both questionnaires, are composed of EDNOS. The diagnostic category EDNOS was proposed and investigated a number of years after the development of the EAT40 and the EDI (Fairburn & Harrison 2003, Norring & Palmer 2005, Walsh & Satir 2005). The EAT was in fact planned to detect possible cases of AN full syndrome and the EDI to describe the psychopathology of subjects

affected by AN and BN. Because of this it is likely that the introduction of the EDNOS diagnoses as part of the eating disorders increases the PPV of the two questionnaires, but lowers their sensitivity. By the way, in a review article about the state of eating disorder screens Jacobi *et al.* wrote that no screen was found explicitly aimed at detection of partial or subclinical ED (Jacobi *et al.* 2004). Furthermore, people affected by ED tend to deny the illness and/or hide it. The great frequency with which the mass media describe ED symptoms and in recent years the appearance of harmful pro-AN chat lines on the web has furthered a positive view of these disorders and probably makes it increasingly easier to hide them.

Finally, some words about the one-point-prevalence of EDNOS: 8.6% in our study and 4.9 % in the Spanish study (considering only the females). How can we explain the difference? Is it possible that the type of diagnostic interview (SCAN vs. EDE) plays a role?

Also in another Spanish study of a non-clinical sample –concerning the validity of the EAT using the DSM-IV (American Psychiatric Association, 1994) and the ICD-10 (World Health Organization 1992) diagnostic criteria for ED– in which almost all 18 year-old subjects (similar to those of our sample) had been interviewed individually, there was a much lower EDNOS prevalence (Canals *et al.* 2002).

We are not able, obviously, to exclude sociocultural factors to explain these differences. On the other hand, they may also be due to the fact that the diagnostic interview used by Canals *et al.* (2002) was the same Spanish version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Vazquez-Baquero *et al.* 1994) that Rodríguez-Cano *et al.* employed in their research

(Rodriguez-Cano *et al.* 2005). Using the SCAN Canals and coll. found only one case of AN and one of BN in a sample of 152 girls (Canals *et al.* 2002). Another two cases of atypical AN were found in this study using diagnostic clinical methods and criteria that were not described.

It is interesting to examine the EAT results. The mean EAT scores (16.6 ± 11.5 in our study vs. 15.7 ± 11.6 in the Spanish study) and the percentages of subjects with a score 30 or more (13.4% vs 9.3%) do not show statistically significant differences. But when comparing the percentages of subjects with EDNOS diagnoses, the differences seem particularly significant (8.6% in our study vs. 1.3% in the Spanish study, $p = 0.007$).

In other words, according to the EAT40 total score, the samples appear to be very similar. It is therefore possible that the difference in the frequency of EDNOS diagnoses depends fundamentally on differences between the two diagnostic interviews used. The Spanish studies did not use a standardised interview constructed specifically for the eating disorders and able to single out EDNOS. The EDE 12.0D, on the other hand, used in the current study, addresses both requirements (Cuzzolaro & Petrilli 1988, Fairburn & Cooper 1993).

Our research has a serious limit i.e. the small sample size. As we said in the section Subjects and Methods, this sample size was the best compromise between the costs and the purposes of the study, that was especially aimed at identifying people suffering from EDNOS. One way of dealing with this is to rely on multi-centred community screening, in which each of the research groups administers to all both the self-report instrument and diagnostic interview. Studies of this type would significantly improve our

knowledge of the epidemiology of eating disorders. One further benefit from such work might be the improvement or development of self-report instruments. As Wilson and Jungner have written (Wilson & Jungner 1968) "The evaluation of screening may be considered from two separate aspects, which yet have a certain connection with each other. These aspects are, firstly, the evaluation of tests or examinations and, secondly, the evaluation of results".

From a psychopathological perspective, through the identification of larger numbers of cases, such studies would permit a better knowledge of features and course of EDNOS and thus increase the probabilities of improving the current classification scheme. In fact, a number of recent studies comparing full- and partial-syndrome ED have indicated that full cases cannot be meaningfully distinguished from their sub-threshold counterparts (e.g. Crow *et al.* 2002, Walsh & Satir 2005). Furthermore, according to Fairburn & Harrison (2003), the current classification "is unsatisfactory and anomalous, in that about half the cases seen in clinical practice are relegated to atypical or not otherwise specified group".

Conclusion

The main finding of this study is that using a two stage screening approach leads to a high rate of false negatives which might lead to a significant underestimation of the prevalence of ED, particularly of EDNOS, too often classified as non eating disordered subjects.

A second benefit is a better knowledge of the psychometric features of two largely used rating scales, EAT and EDI.

References

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, DSM-IV*. Washington DC: American Psychiatric Association; 1994.
- Beglin S, Fairburn C. What is meant by the term "binge"? *Am J Psychiatry* 1992; 149: 123-124.
- Canals J, Carbajo G, Fernandez-Ballart J. Discriminant validity of the Eating Attitudes Test according to American Psychiatric Association and World Health Organization criteria of eating disorders. *Psychol Rep* 2002; 91(3 Pt 2): 1052-1056.
- Cooper Z, Fairburn C. The Eating Disorder Examination. *Int J Eat Disord* 1987; 6: 1-8.
- Crow S, Agras W, Halmi K, Mitchell J, Kraemer H. Full syndromal versus subthreshold anorexia nervosa, bulimia nervosa, and binge eating disorder: a multicenter study. *Int J Eat Disord* 2002; 32: 309-318.
- Cuzzolaro M, Petrilli A. Validazione della versione italiana dell'EAT-40. *Psichiatria dell'Infanzia e dell'Adolescenza* 1988; 55: 209-217.
- Fairburn CG, Cooper Z. The Eating Disorder Examination (12th edition). In: Fairburn CG, Wilson GT, eds. *Binge Eating. Nature Assessment and Treatment*. New York: The Guilford Press; 1993. p 317-360.
- Fairburn CG, Harrison PJ. Eating disorders. *Lancet* 2003; 361(9355): 407-416.
- Garfinkel P, Newman A. The Eating Attitudes Test: Twenty-five years later. *Eat Weight Disord* 2001; 6: 1-24.
- Garner D, Garfinkel P. The Eating Attitudes Test: an index of the symptoms of anorexia nervosa. *Psychol Med* 1979; 9: 273-279.
- Garner D, Olmstead M, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *Int J Eat Disord* 1983; 1: 15-34.
- Garner D. *EDI 2. Eating Disorder Inventory-2. Professional Manual*. Odessa: Psychological Assessment Resources Inc; 1991.
- Jacobi C, Abascal L, Taylor CB. Screening for eating disorders and high-risk behavior: caution. *Int J Eat Disord* 2004; 36(3): 280-295.
- Lee S, Lee AM, Ngai E, Lee DT, Wing YK. Rationales for food refusal in chinese patients with anorexia nervosa. *Int J Eat Disord* 2001; 29(2): 224-229.
- Norring C, Palmer B. *EDNOS: Eating Disorders Not Otherwise Specified. Scientific and clinical perspectives on the other eating disorders*. Hove: Routledge; 2005.
- Ramacciotti C, Dell'Osso L, Paoli R, Ciapparelli A, Kaplan A, Garfinkel P. Characteristics of eating disorder patients without a drive for thinness. *Int J Eat Disord* 2002; 32: 206-212.
- Ricca V, Manucci E, Di Bernardo M, Mezzani B, Carrara S, Rizzello S, et al. Caratteristiche psicopatologiche e cliniche di una popolazione ambulatoriale di pazienti obesi. *Minerva Psichiatr* 1996; 37: 53-58.
- Rodriguez-Cano T, Beato-Fernandez L, Belmonte-Llario A. New contributions to the prevalence of eating disorders in Spanish adolescents: detection of false negatives. *Eur Psychiatry* 2005; 20(2): 173-178.
- Schoemaker C. The principles of screening for Eating Disorders. In: Vandereycken W, Noordenbos G, eds. *The prevention of Eating Disorders*. London: The Athlone Press; 1998; p. 187-213.
- Strober M, Freeman R, Morrell W. Atypical anorexia nervosa: separation from typical cases in course and outcome in a long-term prospective study. *Int J Eat Disord* 1999; 25(2): 135-142.
- Vazquez-Barquero JL, Gaité L, Artal Simon J, Arenal A, Herrera, Castanedo S, et al. Development and verification of the Spanish version of the "scanning system" psychiatric interview. "Questionnaires for clinical evaluation in neuropsychiatry". *Actas Luso Esp Neurol Psiquiatr* 1994; 22: 109-120.
- Walsh B, Satir D. Classification of eating disorders 2002-2003. In: Wonderlich S, Mitchell J, de Zwaan M, Steiger H, eds. *Eating Disorders Review. Part 1*. Oxford: Radcliffe; 2005; p. 155-163.
- Wilson J, Jungner G. *Principles and practice of screening for diseases. Public Health papers, 34*. Genève: World Health Organization; 1968.
- World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders*. London: Gaskell; 1992.

Address for correspondence:
Paul E Garfinkel, T, F
paul.garfinkel@utoronto.ca
CANADA